OM of: US-08-711-417C-165 to: SwissProt_40:* out_format : pfs

Date: Aug 28, 2002 10:21 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/6228611/runat_28082002_100212_13659/app_query.fasta_1.1639
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=rsp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=6228611_@CGN1_1_85 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-08-711-417C-165
Query length: 1551
Database: SwissProt_40:*
Database sequences: 105224
Database length: 38719550
Search time (sec): 44.280000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
SwissProt_40:IKAR_HUMAN	+	2644.50	2637.21	6.5e-140	519	! Q13422 homo sapiens (human).
SwissProt_40:IKAR_MOUSE	+	2426.50	2420.00	8.2e-128	517	! Q03267 mus musculus (mouse).
SwissProt_40:IKAR_CHICK	+	2337.00	2330.79	7.6e-123	518	! O42410 gallus gallus (chicke
SwissProt_40:IKAR_ONCMY	+	1791.50	1787.10	1.4e-92	522	! O13089 oncorhynchus mykiss (
SwissProt_40:HELI_HUMAN	+	1380.00	1376.96	1.0e-69	526	! Q9uks7 homo sapiens (human).
SwissProt_40:HELI_MOUSE	+	1374.50	1371.48	2.0e-69	526	! P81183 mus musculus (mouse).
SwissProt_40:HUNB_CLOAL	+	334.50	335.71	1.1e-11	485	! O96785 clogmia albipunctata
SwissProt_40:ZF37_HUMAN	+	331.50	331.05	1.6e-11	597	! Q9y6q3 homo sapiens (human).
SwissProt_40:ZF37_MOUSE	+	328.50	328.10	2.4e-11	594	! P17141 mus musculus (mouse).
SwissProt_40:ZG20_XENLA	+	323.00	321.58	4.8e-11	675	! P18714 xenopus laevis (afric
SwissProt_40:Z184_HUMAN	+	323.00	320.72	4.8e-11	751	! Q99676 homo sapiens (human).
SwissProt_40:HUNB_TRICA	+	317.00	317.65	1.0e-10	524	! Q01791 tribolium castaneum (
SwissProt_40:YD49_HUMAN	+	315.00	312.88	1.3e-10	739	! Q9p2j8 homo sapiens (human).
SwissProt_40:Y441_HUMAN	+	312.00	310.36	2.0e-10	697	! O43167 homo sapiens (human).
SwissProt_40:HUNB_DROYA	+	311.50	309.18	2.1e-10	759	! O62541 drosophila yakuba (fr
SwissProt_40:ZF90_MOUSE	+	311.00	310.11	2.2e-10	636	! Q61967 mus musculus (mouse).
SwissProt_40:Z33A_HUMAN	+	311.00	308.15	2.2e-10	810	! Q06730 homo sapiens (human).
SwissProt_40:ZN43_HUMAN	+	308.00	305.23	3.3e-10	803	! P17038 homo sapiens (human).
SwissProt_40:ZN85_HUMAN	+	307.00	306.66	3.7e-10	595	! Q03923 homo sapiens (human).
SwissProt_40:HUNB_DROSE	+	306.50	304.21	4.0e-10	757	! O62538 drosophila sechellia
SwissProt_40:HUNB_DROME	+	305.50	303.21	4.5e-10	758	! P05084 drosophila melanogast
SwissProt_40:ZN91_HUMAN	+	305.50	299.56	4.6e-10	1191	! Q05481 homo sapiens (human)
SwissProt_40:ZF29_MOUSE	+	305.00	304.41	4.8e-10	614	! Q07230 mus musculus (mouse).
SwissProt_40:ZN35_HUMAN	+	304.50	305.27	5.0e-10	519	! P13682 homo sapiens (human).
SwissProt_40:Z197_HUMAN	+	304.00	299.24	5.5e-10	1029	! O14709 homo sapiens (human)
SwissProt_40:ZN84_HUMAN	+	303.50	301.43	5.8e-10	738	! P51523 homo sapiens (human).
SwissProt_40:Z337_HUMAN	+	303.00	300.79	6.2e-10	751	! Q9y3m9 homo sapiens (human).
SwissProt_40:HUNB_MUSDO	+	302.00	299.11	7.0e-10	817	! Q01778 musca domestica (hous
SwissProt_40:ZN83_HUMAN	+	301.00	303.34	7.8e-10	428	! P51522 homo sapiens (human).
SwissProt_40:Z287_MOUSE	+	301.00	298.71	8.0e-10	759	! Q9eqb9 mus musculus (mouse).
SwissProt_40:ZF95_HUMAN	+	301.00	297.90	8.0e-10	839	! Q9y218 homo sapiens (human).
SwissProt_40:ZF95_MOUSE	+	300.00	297.10	9.1e-10	819	! Q9z1d8 mus musculus (mouse).
SwissProt_40:Z189_HUMAN	+	299.50	298.77	9.6e-10	626	! O75820 homo sapiens (human).
SwissProt_40:ZN42_HUMAN	+	299.50	297.49	9.7e-10	734	! P28698 homo sapiens (human).
SwissProt_40:TC17_MOUSE	+	298.50	298.51	1.1e-09	572	! Q61751 mus musculus (mouse).
SwissProt_40:ZF35_MOUSE	+	298.50	298.39	1.1e-09	580	! P15620 mus musculus (mouse).
SwissProt_40:ZN44_HUMAN	+	298.50	298.27	1.1e-09	589	! P15621 homo sapiens (human).
SwissProt_40:Z135_HUMAN	+	298.00	299.61	1.2e-09	469	! P52742 homo sapiens (human).
SwissProt_40:Z225_HUMAN	+	297.00	295.31	1.3e-09	706	! Q9uk10 homo sapiens (human).
SwissProt_40:XFIN_XENLA	+	295.50	288.58	1.6e-09	1350	! P08045 xenopus laevis (afri
SwissProt_40:Z142_HUMAN	+	295.50	286.78	1.7e-09	1687	! P52746 homo sapiens (hum
SwissProt_40:ZO28_XENLA	+	294.50	296.66	1.8e-09	439	! P18747 xenopus laevis (af
SwissProt_40:Z157_HUMAN	+	294.50	295.51	1.8e-09	506	! P51786 homo sapiens (huma
SwissProt_40:ZN74_HUMAN	+	294.00	293.08	1.9e-09	643	! Q16587 homo sapiens (huma
SwissProt_40:HUNB_DROOR	+	293.50	291.15	2.1e-09	767	! O62537 drosophila orena (

seq_name: SwissProt_40:IKAR_HUMAN

seq_documentation_block:
ID IKAR_HUMAN STANDARD; PRT; 519 AA.
AC Q13422; O00598;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-binding protein Ikaros (Lymphoid transcription factor LyF-1).
GN ZNFN1A1 OR IKAROS OR IK1 OR LYF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Bone marrow;
RX MEDLINE=96252222; PubMed=8964602;
RA Nietfeld W., Meyerhans A.;
RT "Cloning and sequencing of hIk-1, a cDNA encoding a human homologue
RT of mouse Ikaros/LyF-1.";
RL Immunol. Lett. 49:139-141(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96132984; PubMed=8543809;
RA Molnar A., Wu P., Largespada D.A., Vortkamp A., Scherer S.,
RA Copeland N.G., Jenkins N.A., Bruns G., Georgopoulos K.;
RT "The Ikaros gene encodes a family of lymphocyte-restricted zinc
RT finger DNA binding proteins, highly conserved in human and mouse.";
RL J. Immunol. 156:585-592(1996).
CC -!- FUNCTION: BINDS AND ACTIVATES THE ENHANCER (DELTA-A ELEMENT) OF
CC THE CD3-DELTA GENE. FUNCTIONS IN THE SPECIFICATION AND THE
CC MATURATION OF THE T LYMPHOCYTE. ALSO INTERACTS WITH A CRITICAL
CC CONTROL ELEMENT IN THE TDT (TERMINAL DEOXYNUCLEOTIDYLTRANSFERASE)
CC PROMOTER AS WELL AS WITH THE PROMOTERS FOR OTHER GENES EXPRESSED
CC DURING EARLY STAGES OF B AND T CELL DEVELOPMENT.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS: 6 ISOFORMS; IK1 (SHOWN HERE), IK2, IK3, IK4,
CC IK5 AND IK6; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: ABUNDANTLY EXPRESSED IN THYMUS, SPLEEN AND
CC PERIPHERAL BLOOD LEUKOCYTES AND LYMPH NODES. LOWER EXPRESSION IN
CC BONE MARROW AND SMALL INTESTINE.
CC -!- SIMILARITY: BELONGS TO THE IKAROS FAMILY OF ZINC-FINGER PROTEINS.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; U40462; AAC50459.1; -.
DR EMBL; S80876; AAB50683.1; -.
DR HSSP; P08047; 1SP2.
DR TRANSFAC; T02702; -.
DR MIM; 603023; -.
DR InterPro; IPR000822; Znf-C2H2.
DR Pfam; PF00096; zf-C2H2; 5.
DR PRINTS; PR00048; ZINCFINGER.
DR SMART; SM00355; ZnF_C2H2; 6.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW DNA-binding; Nuclear protein; Repeat; Alternative splicing.
FT DOMAIN 117 224 ZINC-FINGERS I.
FT ZN_FING 117 139 C2H2-TYPE.

```
FT   ZN_FING     145    167       C2H2-TYPE.
FT   ZN_FING     173    195       C2H2-TYPE.
FT   ZN_FING     201    224       C2H2-TYPE.
FT   DOMAIN      462    514       ZINC-FINGERS II.
FT   ZN_FING     462    484       C2H2-TYPE.
FT   ZN_FING     490    514       C2H2-TYPE.
FT   VARSPLIC     10     53       MISSING (IN ISOFORM IK4).
FT   VARSPLIC     54    140       MISSING (IN ISOFORM IK2).
FT   VARSPLIC     54    283       MISSING (IN ISOFORM IK6).
FT   VARSPLIC    197    283       MISSING (IN ISOFORM IK3 AND ISOFORM IK4).
FT   VARSPLIC    141    283       MISSING (IN ISOFORM IK5).
FT   CONFLICT     11     12       QV -> FS (IN REF. 2).
FT   CONFLICT    214    214       S -> T (IN REF. 2).
FT   CONFLICT    245    245       N -> K (IN REF. 2).
FT   CONFLICT    296    296       MISSING (IN REF. 2).
FT   CONFLICT    298    298       S -> T (IN REF. 2).
FT   CONFLICT    352    355       KPLA -> RRS (IN REF. 2).
FT   CONFLICT    372    372       N -> Y (IN REF. 2).
FT   CONFLICT    420    426       PHARNGL -> RRAQRV (IN REF. 2).
SQ   SEQUENCE   519 AA;  57528 MW;  7B0129C4E3FE41A8 CRC64;


alignment_scores:
            Quality: 2644.50              Length:    519
              Ratio:  5.175                Gaps:      3
 Percent Similarity:  98.459   Percent Identity: 96.724

alignment_block:
 US-08-711-417C-165 x IKAR_HUMAN   ..

 Align seg 1/1  to: IKAR_HUMAN  from: 1  to: 519

       1 ATGGATGCTGACGAGGGTCAAGACATGTCTTTCTCATCAGGGAAGGAAAG 50
         ||||||||||||||||||||||||||||||      ||||||||||||||
       1 MetAspAlaAspGluGlyGlnAspMetSerGlnValSerGlyLysGluSe 17

      51 CCCCCCTGTAAGCGATACTCCAGATGAGGGCGATGAGCCCATGCCGATCC 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      17 rProProValSerAspThrProAspGluGlyAspGluProMetProIleP 34

     101 CCGAGGACCTCTCCACCACCTCGGGAGGACAGCAAAGCTCCAAGAGTGAC 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      34 roGluAspLeuSerThrThrSerGlyGlyGlnGlnSerSerLysSerAsp 50

     151 AGAGTCGTGGCCAGTAATGTTAAAGTAGAGACTCAGAGTGATGAAGAGAA 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ArgValValAlaSerAsnValLysValGluThrGlnSerAspGluGluAs 67

     201 TGGGCGTGCCTGTGAAATGAATGGGGAAGAATGTGCGGAGGATTTACGAA 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      67 nGlyArgAlaCysGluMetAsnGlyGluGluCysAlaGluAspLeuArgM 84

     251 TGCTTGATGCCTCGGGAGAGAAAATGAATGGCTCCCACAGGGACCAAGGC 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      84 etLeuAspAlaSerGlyGluLysMetAsnGlySerHisArgAspGlnGly 100

     301 AGCTCGGCTTTGTCGGGAGTTGGAGGCATTCGACTTCCTAACGGAAAACT 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     101 SerSerAlaLeuSerGlyValGlyGlyIleArgLeuProAsnGlyLysLe 117

     351 AAAGTGTGATATCTGTGGGATCATTTGCATCGGGCCCAATGTGCTCATGG 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     117 uLysCysAspIleCysGlyIleIleCysIleGlyProAsnValLeuMetV 134

     401 TTCACAAAAGAAGCCACACTGGAGAACGGCCCTTCCAGTGCAATCAGTGC 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     134 alHisLysArgSerHisThrGlyGluArgProPheGlnCysAsnGlnCys 150

     451 GGGGCCTCATTCACCCAGAAGGGCAACCTGCTCCGGCACATCAAGCTGCA 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GlyAlaSerPheThrGlnLysGlyAsnLeuLeuArgHisIleLysLeuHi 167

     501 TTCCGGGGAGAAGCCCTTCAAATGCCACCTCTGCAACTACGCCTGCCGCC 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     167 sSerGlyGluLysProPheLysCysHisLeuCysAsnTyrAlaCysArgA 184

     551 GGAGGGACGCCCTCACTGGCCACCTGAGGACGCACTCCGTTGGTAAACCT 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     184 rgArgAspAlaLeuThrGlyHisLeuArgThrHisSerValGlyLysPro 200

     601 CACAAATGTGGATATTGTGGCCGAAGCTATAAACAGCGAACGTCTTTAGA 650
         ||||||||||||||||||||||||||||||||||||:::|||||||||||
     201 HisLysCysGlyTyrCysGlyArgSerTyrLysGlnArgSerSerLeuGl 217

     651 GGAACATAAAGAGCGCTGCCACAACTACTTGGAAAGCATGGGCCTTCCGG 700
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     217 uGluHisLysGluArgCysHisAsnTyrLeuGluSerMetGlyLeuProG 234

     701 GCACACTGTACCCAGTCATTAAAGAAGAAACTAAGCACAGTGAAATGGCA 750
         ||||||||||||||||||||||||||||||:::|||||||||||||||||
     234 lyThrLeuTyrProValIleLysGluGluThrAsnHisSerGluMetAla 250

     751 GAAGACCTGTGCAAGATAGGATCAGAGAGATCTCTCGTGCTGGACAGACT 800
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GluAspLeuCysLysIleGlySerGluArgSerLeuValLeuAspArgLe 267

     801 AGCAAGTAATGTCGCCAAACGTAAGAGCTCTATGCCTCAGAAATTTCTTG 850
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     267 uAlaSerAsnValAlaLysArgLysSerSerMetProGlnLysPheLeuG 284

     851 GGGACAAGGGCCTGTCCGACACGCCCTACGAC...AGTGCCACGTACGAG 897
         |||||||||||||||||||||||||||||||   ||||||:::||||||
     284 lyAspLysGlyLeuSerAspThrProTyrAspSerSerAlaSerTyrGlu 300

     898 AAGGAGAACGAAATGATGAAGTCCCACGTGATGGACCAAGCCATCAACAA 947
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     301 LysGluAsnGluMetMetLysSerHisValMetAspGlnAlaIleAsnAs 317

     948 CGCCATCAACTACCTGGGGGCCGAGTCCCTGCGCCCGCTGGTGCAGACGC 997
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     317 nAlaIleAsnTyrLeuGlyAlaGluSerLeuArgProLeuValGlnThrP 334

     998 CCCCGGGCGGTTCCGAGGTGGTCCCGGTCATCAGCCCGATGTACCAGCTG 1047
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     334 roProGlyGlySerGluValValProValIleSerProMetTyrGlnLeu 350

    1048 CACAGG...CGCTCGGAGGGCACCCCGCGCTCCAACCACTCGGCCCAGGA 1094
         |||:::      :::|||||||||||||||||||||||||||||||||||
     351 HisLysProLeuAlaGluGlyThrProArgSerAsnHisSerAlaGlnAs 367

    1095 CAGCGCCGTGGAGTACCTGCTGCTGCTCTCCAAGGCCAAGTTGGTGCCCT 1144
         |||||||||||||   ||||||||||||||||||||||||||||||||||
     367 pSerAlaValGluAsnLeuLeuLeuLeuSerLysAlaLysLeuValProS 384

    1145 CGGAGCGCGAGGCGTCCCCGAGCAACAGCTGCCAAGACTCCACGGACACC 1194
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     384 erGluArgGluAlaSerProSerAsnSerCysGlnAspSerThrAspThr 400

    1195 GAGAGCAACAACGAGGAGCAGCGCAGCGGTCTTATCTACCTGACCAACCA 1244
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     401 GluSerAsnAsnGluGluGlnArgSerGlyLeuIleTyrLeuThrAsnHi 417

    1245 CATCGCCCGACGCGCGCAACGC...GTGTCGCTCAAGGAGGAGCACCGCG 1291
         |||||||   :::|||::::::   :::|||||||||||||||||||||
     417 sIleAlaProHisAlaArgAsnGlyLeuSerLeuLysGluGluHisArgA 434

    1292 CCTACGACCTGCTGCGCGCCGCCTCCGAGAACTCGCAGGACGCGCTCCGC 1341
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     434 laTyrAspLeuLeuArgAlaAlaSerGluAsnSerGlnAspAlaLeuArg 450

    1342 GTGGTCAGCACCAGCGGGGAGCAGATGAAGGTGTACAAGTGCGAACACTG 1391
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     451 ValValSerThrSerGlyGluGlnMetLysValTyrLysCysGluHisCy 467

    1392 CCGGGTGCTCTTCCTGGATCACGTCATGTACACCATCCACATGGGCTGCC 1441
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     467 sArgValLeuPheLeuAspHisValMetTyrThrIleHisMetGlyCysH 484

    1442 ACGGCTTCCGTGATCCTTTTGAGTGCAACATGTGCGGCTACCACAGCCAG 1491
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     484 isGlyPheArgAspProPheGluCysAsnMetCysGlyTyrHisSerGln 500

    1492 GACCGGTACGAGTTCTCGTCGCACATAACGCGAGGGGAGCACCGCTTCCA 1541
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 AspArgTyrGluPheSerSerHisIleThrArgGlyGluHisArgPheHi 517

    1542 CATGAGC 1548
         |||||||
     517 sMetSer 519

seq_name: SwissProt_40:IKAR_MOUSE

seq_documentation_block:
ID   IKAR_MOUSE     STANDARD;      PRT;   517 AA.
AC   Q03267; Q64044; Q64045; Q64051;
DT   01-OCT-1993 (Rel. 27, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA-binding protein Ikaros (Lymphoid transcription factor LyF-1).
GN   ZNFN1A1 OR IKAROS OR LYF1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM V).
RC   TISSUE=Embryo;
RX   MEDLINE=93068267; PubMed=1439790;
RA   Georgopoulos K., Moore D.D., Derfler B.;
RT   "Ikaros, an early lymphoid-specific transcription factor and a
RT   putative mediator for T cell commitment.";
RL   Science 258:808-812(1992).
RN   [2]
RP   SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ALTERNATIVE SPLICING.
RX   MEDLINE=95021239; PubMed=7935426;
RA   Hahm K., Ernst P., Lo K., Kim G.S., Turck C., Smale S.T.;
RT   "The lymphoid transcription factor LyF-1 is encoded by specific,
RT   alternatively spliced mRNAs derived from the Ikaros gene.";
RL   Mol. Cell. Biol. 14:7111-7123(1994).
CC   -!- FUNCTION: BINDS AND ACTIVATES THE ENHANCER (DELTA-A ELEMENT) OF
CC       THE CD3-DELTA GENE. FUNCTIONS IN THE SPECIFICATION AND THE
CC       MATURATION OF THE T LYMPHOCYTE. ALSO INTERACTS WITH A CRITICAL
CC       CONTROL ELEMENT IN THE TDT (TERMINAL DEOXYNUCLEOTIDYLTRANSFERASE)
CC       PROMOTER AS WELL AS WITH THE PROMOTERS FOR OTHER GENES EXPRESSED
CC       DURING EARLY STAGES OF B AND T CELL DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS: 6 ISOFORMS; I, II, III, IV, V AND VI (SHOWN
CC       HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN T-CELLS AND THEIR
CC       PROGENITORS, AND ALSO IN B-CELLS.
CC   -!- SIMILARITY: BELONGS TO THE IKAROS FAMILY OF ZINC-FINGER PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L03547; AAA66193.1; -.
DR   EMBL; S74517; AAB32248.2; ALT_SEQ.
DR   EMBL; S74518; AAB32249.2; -.
DR   EMBL; S74708; AAB32250.2; -.
DR   HSSP; P08047; 1SP2.
DR   TRANSFAC; T01470; -.
DR   MGD; MGI:1342540; Znfn1a1.
DR   InterPro; IPR000822; Znf-C2H2.
DR   Pfam; PF00096; zf-C2H2; 5.
DR   PRINTS; PR00048; ZINCFINGER.
DR   SMART; SM00355; ZnF_C2H2; 6.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.
KW   Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW   DNA-binding; Nuclear protein; Repeat; Alternative splicing.
FT   DOMAIN       117    223       ZINC-FINGERS I.
FT   ZN_FING      117    139       C2H2-TYPE.
FT   ZN_FING      144    166       C2H2-TYPE.
FT   ZN_FING      172    194       C2H2-TYPE.
FT   ZN_FING      200    223       C2H2-TYPE.
FT   DOMAIN       457    479       ZINC-FINGERS II.
FT   ZN_FING      457    479       C2H2-TYPE.
FT   VARSPLIC      53     53       M -> VAYGADGFRDFHAIISDRGM (IN ISOFORM II
FT                                 AND ISOFORM IV).
FT   VARSPLIC      54    140       MISSING (IN ISOFORM V).
FT   VARSPLIC      54    282       MISSING (IN ISOFORM I AND ISOFORM II).
FT   VARSPLIC     141    282       MISSING (IN ISOFORM III AND ISOFORM IV).
FT   CONFLICT     234    235       VC -> MY (IN REF. 2).
FT   CONFLICT     480    482       MISSING (IN REF. 2).
SQ   SEQUENCE   517 AA;  57336 MW;  1052B8E76AF24287 CRC64;


alignment_scores:
            Quality: 2426.50              Length:    521
              Ratio:  4.892                Gaps:      7
 Percent Similarity:  95.202   Percent Identity: 89.443

alignment_block:
 US-08-711-417C-165 x IKAR_MOUSE   ..

 Align seg 1/1  to: IKAR_MOUSE  from: 1  to: 517

       1 ATGGATGCTGACGAGGGTCAAGACATGTCTTTCTCATCAGGGAAGGAAAG 50
         ||||||:::|||||||||||||||||||||      ||||||||||||||
       1 MetAspValAspGluGlyGlnAspMetSerGlnValSerGlyLysGluSe 17

      51 CCCCCCTGTAAGCGATACTCCAGATGAGGGCGATGAGCCCATGCCGATCC 100
         ||||||||||||||||||||||||||||||||||||||||||||||:::|
      17 rProProValSerAspThrProAspGluGlyAspGluProMetProValP 34

     101 CCGAGGACCTCTCCACCACCTCGGGAGGACAGCAAAGCTCCAAGAGTGAC 150
         ||||||||||||||||||||||||||:::||||||:::||||||||||||
      34 roGluAspLeuSerThrThrSerGlyAlaGlnGlnAsnSerLysSerAsp 50

     151 AGAGTCGTGGCCAGTAATGTTAAAGTAGAGACTCAGAGTGATGAAGAGAA 200
         |||   ::::::||||||||||||||||||||||||||||||||||||||
      51 ArgGlyMetGlySerAsnValLysValGluThrGlnSerAspGluGluAs 67

     201 TGGGCGTGCCTGTGAAATGAATGGGGAAGAATGTGCGGAGGATTTACGAA 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 nGlyArgAlaCysGluMetAsnGlyGluGluCysAlaGluAspLeuArgM 84

     251 TGCTTGATGCCTCGGGAGAGAAAATGAATGGCTCCCACAGGGACCAAGGC 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 etLeuAspAlaSerGlyGluLysMetAsnGlySerHisArgAspGlnGly 100

     301 AGCTCGGCTTTGTCGGGAGTTGGAGGCATTCGACTTCCTAACGGAAAACT 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 SerSerAlaLeuSerGlyValGlyGlyIleArgLeuProAsnGlyLysLe 117

     351 AAAGTGTGATATCTGTGGGATCATTTGCATCGGGCCCAATGTGCTCATGG 400
         ||||||||||||||||||:::||||||||||||||||||||||||||||
     117 uLysCysAspIleCysGlyIleValCysIleGlyProAsnValLeuMetV 134

     401 TTCACAAAAGAAGCCACACTGGAGAACGGCCCTTCCAGTGCAATCAGTGC 450
         ||||||||||||||||||   ||||||||||||||||||||||||||||
     134 alHisLysArgSerHisThr...GluArgProPheGlnCysAsnGlnCys 149

     451 GGGGCCTCATTCACCCAGAAGGGCAACCTGCTCCGGCACATCAAGCTGCA 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     150 GlyAlaSerPheThrGlnLysGlyAsnLeuLeuArgHisIleLysLeuHi 166
```

```
 501 TTCCGGGGAGAAGCCCTTCAAATGCCACCTCTGCAACTACGCCTGCCGCC 550
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 166 sSerGlyGluLysProPheLysCysHisLeuCysAsnTyrAlaCysArgA 183

 551 GGAGGGACGCCCTCACTGGCCACCTGAGGACGCACTCCGTTGGTAAACCT 600
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 183 rgArgAspAlaLeuThrGlyHisLeuArgThrHisSerValGlyLysPro 199

 601 CACAAATGTGGATATTGTGGCCGAAGCTATAAACAGCGAACGTCTTTAGA 650
     ||||||||||||||||||||||||||||||||||||:::|||||||||
 200 HisLysCysGlyTyrCysGlyArgSerTyrLysGlnArgSerSerLeuGl 216

 651 GGAACATAAAGAGCGCTGCCACAACTACTTGGAAAGCATGGGCCTTCCGG 700
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 216 uGluHisLysGluArgCysHisAsnTyrLeuGluSerMetGlyLeuProG 233

 701 GCACACTGTACCCAGTCATTAAAGAAGAAACTAAGCACAGTGAAATGGCA 750
     ||   :::   |||||||||||||||||||||:::|||:::|||||||||
 233 ly...ValCysProValIleLysGluGluThrAsnHisAsnGluMetAla 248

 751 GAAGACCTGTGCAAGATAGGATCAGAGAGATCTCTCGTGCTGGACAGACT 800
     ||||||||||||||||||||:::|||||||||||||||||||||||||||
 249 GluAspLeuCysLysIleGlyAlaGluArgSerLeuValLeuAspArgLe 265

 801 AGCAAGTAATGTCGCCAAACGTAAGAGCTCTATGCCTCAGAAATTTCTTG 850
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 265 uAlaSerAsnValAlaLysArgLysSerSerMetProGlnLysPheLeuG 282

 851 GGGACAAGGGCCTGTCCGACACGCCCTACGACAGTGCCACGTACGAGAAG 900
     ||||||||   |||||||||   ||||||||||||||:::|||||||||
 282 lyAspLysCysLeuSerAspMetProTyrAspSerAlaAsnTyrGluLys 298

 901 GAGAACGAAATGATGAAGTCCCACGTGATGGACCAAGCCATCAACAACGC 950
     |||   :::||||||   ||||||||||||||||||||||||||||||||
 299 Glu...AspMetMetThrSerHisValMetAspGlnAlaIleAsnAsnAl 314

 951 CATCAACTACCTGGGGGCCGAGTCCCTGCGCCCGCTGGTGCAGACGCCCC 1000
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 314 aIleAsnTyrLeuGlyAlaGluSerLeuArgProLeuValGlnThrProP 331

1001 CGGGCGGTTCCGAGGTGGTCCCGGTCATCAGCCCGATGTACCAGCTGCAC 1050
     ||||||:::|||||||||||||||||||||||   ||||||||||||||||
 331 roGlySerSerGluValValProValIleSerSerMetTyrGlnLeuHis 347

1051 AGG...CGCTCGGAGGGCACCCCGCGCTCCAACCACTCGGCCCAGGACAG 1097
     :::      |||:::|||   |||||||||||||||||||||||||||||
 348 LysProProSerAspGlyProProArgSerAsnHisSerAlaGlnAsp.. 363

1098 CGCCGTGGAGTACCTGCTGCTGCTCTCCAAGGCCAAGTTGGTGCCCTCGG 1147
      ||||||:::   |||||||||||||||||||||||   |||   ||||
  364 .AlaValAspAsnLeuLeuLeuLeuSerLysAlaLysSerValSerSerG 380

1148 AGCGCGAGGCGTCCCCGAGCAACAGCTGCCAAGACTCCACGGACACCGAG 1197
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  380 luArgGluAlaSerProSerAsnSerCysGlnAspSerThrAspThrGlu 396

1198 AGCAACAACGAGGAGCAGCGCAGCGGTCTTATCTACCTGACCAACCACAT 1247
     ||||||   |||||||||||||||||||||||||||||||||||||||||
  397 SerAsnAlaGluGluGlnArgSerGlyLeuIleTyrLeuThrAsnHisIl 413

1248 CGCCCGACGCGCGCAACGC...GTGTCGCTCAAGGAGGAGCACCGCGCCT 1294
     |      :::|||:::::   ::::::|||||||||||:::|||||||
  413 eAsnProHisAlaArgAsnGlyLeuAlaLeuLysGluGluGlnArgAlaT 430

1295 ACGACCTGCTGCGCGCCGCCTCCGAGAACTCGCAGGACGCGCTCCGCGTG 1344
     ||::::::|||||||||||||||||||||||||||||||:::||||||
  430 yrGluValLeuArgAlaAlaSerGluAsnSerGlnAspAlaPheArgVal 446

1345 GTCAGCACCAGCGGGGAGCAGATGAAGGTGTACAAGTGCGAACACTGCCG 1394
     ||||||||||||||||||:::|||||||||||||||||||||||||||||
  447 ValSerThrSerGlyGluGlnLeuLysValTyrLysCysGluHisCysAr 463

1395 GGTGCTCTTCCTGGATCACGTCATGTACACCATCCACATG.....G 1435
```

```
     |||||||||||||||||||||||||||||||||||||||||||||||||
  463 gValLeuPheLeuAspHisValMetTyrThrIleHisMetGlyCysHisG 480

 1436 GCTGCCACGGCTTCCGTGATCCTTTTGAGTGCAACATGTGCGGCTACCAC 1485
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  480 lyCysHisGlyPheArgAspProPheGluCysAsnMetCysGlyTyrHis 496

 1486 AGCCAGGACCGGTACGAGTTCTCGTCGCACATAACGCGAGGGGAGCACCG 1535
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  497 SerGlnAspArgTyrGluPheSerSerHisIleThrArgGlyGluHisAr 513

 1536 CTTCCACATGAGC 1548
      |:::|||:::|||
  513 gTyrHisLeuSer 517
```

```
seq_name: SwissProt_40:IKAR_CHICK

seq_documentation_block:
ID   IKAR_CHICK     STANDARD;      PRT;   518 AA.
AC   O42410;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA-binding protein Ikaros.
GN   IKAROS OR IK.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H.B2; TISSUE=Thymus;
RX   MEDLINE=97439462; PubMed=9295018;
RA   Liippo J., Lassila O.;
RT   "Avian Ikaros gene is expressed early in embryogenesis.";
RL   Eur. J. Immunol. 27:1853-1857(1997).
CC   -!- FUNCTION: BINDS AND ACTIVATES THE ENHANCER (DELTA-A ELEMENT) OF
CC       THE CD3-DELTA GENE. FUNCTIONS IN THE SPECIFICATION AND THE
CC       MATURATION OF THE T LYMPHOCYTE. ALSO INTERACTS WITH A CRITICAL
CC       CONTROL ELEMENT IN THE TDT (TERMINAL DEOXYNUCLEOTIDYLTRANSFERASE)
CC       PROMOTER AS WELL AS WITH THE PROMOTERS FOR OTHER GENES
CC       EXPRESSED DURING EARLY STAGES OF B AND T CELL DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS: A NUMBER OF ISOFORMS ARE PRODUCED BY
CC       ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN EMBRYONIC HEMATOPOIETIC ORGANS
CC       SUCH AS THE BURSA OF FABRICIUS, THYMUS AND SPLEEN. IN THE ADULT,
CC       EXPRESSED IN SPLEEN, THYMUS, BURSA AND PERIPHERAL BLOOD
CC       LEUKOCYTES.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN EMBRYO FROM EMBRYONIC DAY
CC       2 ONWARDS.
CC   -!- SIMILARITY: BELONGS TO THE IKAROS FAMILY OF ZINC-FINGER PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y11833; CAA72531.1; -.
DR   HSSP; P15822; 1BBO.
DR   InterPro; IPR000822; Znf-C2H2.
DR   Pfam; PF00096; zf-C2H2; 5.
DR   PRINTS; PR00048; ZINCFINGER.
DR   SMART; SM00355; ZnF_C2H2; 6.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW   DNA-binding; Nuclear protein; Repeat; Alternative splicing.
FT   DOMAIN      117    224       ZINC-FINGERS I.
```

```
FT   ZN_FING     117    139       C2H2-TYPE.
FT   ZN_FING     145    167       C2H2-TYPE.
FT   ZN_FING     173    195       C2H2-TYPE.
FT   ZN_FING     201    224       C2H2-TYPE.
FT   DOMAIN      461    513       ZINC-FINGERS II.
FT   ZN_FING     461    483       C2H2-TYPE.
FT   ZN_FING     489    513       C2H2-TYPE.
SQ   SEQUENCE   518 AA;  57586 MW;  4DF15B677E616135 CRC64;


alignment_scores:
           Quality: 2337.00              Length:    519
             Ratio:  4.712                Gaps:        4
 Percent Similarity:  95.568   Percent Identity: 83.622

alignment_block:
 US-08-711-417C-165 x IKAR_CHICK   ..

 Align seg 1/1  to: IKAR_CHICK  from: 1  to: 518

       1 ATGGATGCTGACGAGGGTCAAGACATGTCTTTCTCATCAGGGAAGGAAAG 50
         |||:::::::|||||:::|||||||||||||      ||||||||||||||
       1 MetGluThrAspGluAlaGlnAspMetSerGlnValSerGlyLysGluSe 17

      51 CCCCCCTGTAAGCGATACTCCAGATGAGGGCGATGAGCCCATGCCGATCC 100
         |||||||:::|||||||:::|||||||:::::|||||||||||||||:::|
      17 rProProIleSerAspValProAspAspAlaAspGluProMetProValP 34

     101 CCGAGGACCTCTCCACCACCTCGGGAGGACAGCAAAGCTCCAAGAGTGAC 150
         ||||||||||||||||||:::|||||||||||||||   |||::::::
      34 roGluAspLeuSerThrThrThrGlyGlyGlnGlnSerValLysAsnGlu 50

     151 AGAGTCGTGGCCAGTAATGTTAAAGTAGAGACTCAGAGTGATGAAGAGAA 200
         ||||||:::|||:::|||:::|||:::|||||||||||||||||||||||
      51 ArgValLeuAlaGlyAsnIleLysIleGluThrGlnSerAspGluGluAs 67

     201 TGGGCGTGCCTGTGAAATGAATGGGGAAGAATGTGCCGGAGGATTTACGAA 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 nGlyArgAlaCysGluMetAsnGlyGluGluCysAlaGluAspLeuArgM 84

     251 TGCTTGATGCCTCGGGAGAGAAAATGAATGGCTCCCACAGGGACCAAGGC 300
         ||||||||||||||||:::|||||||||||||||:::      |||
      84 etLeuAspAlaSerGlyAspLysMetAsnGlySerHisAsnGlyProGly 100

     301 AGCTCGGCTTTGTCGGGAGTTGGAGGCATTCGACTTCCTAACGGAAAACT 350
         |||:::|||:::|||||||||||||||||||||||||||||||||||||
     101 SerLysAlaMetSerGlyValGlyGlyIleArgLeuProAsnGlyLysLe 117

     351 AAAGTGTGATATCTGTGGGATCATTTGCATCGGGCCCAATGTGCTCATGG 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     117 uLysCysAspIleCysGlyIleIleCysIleGlyProAsnValLeuMetV 134

     401 TTCACAAAAGAAGCCACACTGGAGAACGGCCCTTCCAGTGCAATCAGTGC 450
         |||||:::|||||||||||||||||||||||||||||||||||||||||||
     134 alHisAsnArgSerHisThrGlyGluArgProPheGlnCysAsnGlnCys 150

     451 GGGGCCTCATTCACCCAGAAGGGCAACCTGCTCCGGCACATCAAGCTGCA 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GlyAlaSerPheThrGlnLysGlyAsnLeuLeuArgHisIleLysLeuHi 167

     501 TTCCGGGGAGAAGCCCTTCAAATGCCACCTCTGCAACTACGCCTGCCGCC 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     167 sSerGlyGluLysProPheLysCysHisLeuCysAsnTyrAlaCysArgA 184

     551 GGAGGGACGCCCTCACTGGCCACCTGAGGACGCACTCCGTTGGTAAACCT 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 rgArgAspAlaLeuThrGlyHisLeuArgThrHisSerValGlyLysPro 200

     601 CACAAATGTGGATATTGTGGCCGAAGCTATAAACAGCGAACGTCTTTAGA 650
         ||||||||||||||||||||||||||||||||||:::|||||||||
     201 HisLysCysGlyTyrCysGlyArgSerTyrLysGlnArgSerSerLeuGl 217

     651 GGAACATAAAGAGCGCTGCCACAACTACTTGGAAAGCATGGGCCTTCCGG 700
         |||||||||||||||||||||||||||||||||:::::|||:::::   :
     217 uGluHisLysGluArgCysHisAsnTyrLeuGlnThrMetSerIleSerS 234

     701 GCACACTGTACCCAGTCATTAAAGAAGAAACTAAGCACAGTGAAATGGCA 750
         :::::||||||   ||||||||||||||||||:::::|||||||||||||
     234 erAsnLeuTyrSerValIleLysGluGluThrAsnGlnSerGluMetAla 250

     751 GAAGACCTGTGCAAGATAGGATCAGAGAGATCTCTCGTGCTGGACAGACT 800
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GluAspLeuCysLysIleGlySerGluArgSerLeuValLeuAspArgLe 267

     801 AGCAAGTAATGTCGCCAAACGTAAGAGCTCTATGCCTCAGAAATTTCTTG 850
         ||||||||||||||||||||||||||||||||||||||||||||||:::|
     267 uAlaSerAsnValAlaLysArgLysSerSerMetProGlnLysPheValG 284

     851 GGGACAAGGGCCTGTCCGACACGCCCTACGACAGTGCCACG...TACGAG 897
         ||:::|||   |||||||||   |||||||||:::::|||   ||||||
     284 lyGluLysCysLeuSerAspLeuProTyrAspAlaThrThrAsnTyrGlu 300

     898 AAGGAGAACGAAATGATGAAGTCCCACGTGATGGACCAAGCCATCAACAA 947
         ||||||||||||:::|||:::::::|||||||:::|||||||||||||||
     301 LysGluAsnGluIleMetGlnThrHisValIleAspGlnAlaIleAsnAs 317

     948 CGCCATCAACTACCTGGGGGCCGAGTCCCTGCGCCCGCTGGTGCAGACGC 997
         |||||||:::|||||||||||||||||||||||||||||||||||||||
     317 nAlaIleSerTyrLeuGlyAlaGluSerLeuArgProLeuValGlnThrP 334

     998 CCCCGGGCGGTTCCGAGGTGGTCCCGGTCATCAGCCCGATGTACCAGCTG 1047
         |||||   ||||||||||||||||||||||||||||||||||||||||||
     334 roProValGlySerGluValValProValIleSerProMetTyrGlnLeu 350

    1048 CACAGG...CGCTCGGAGGGCACCCCGCGCTCCAACCACTCGGCCCAGGA 1094
         |||:::      :::::::::      ||||||||||||:::|||||||||
     351 HisLysProLeuGlyAspAsnGlnThrArgSerAsnHisThrAlaGlnAs 367

    1095 CAGCGCCGTGGAGTACCTGCTGCTGCTCTCCAAGGCCAAGTTGGTGCCCT 1144
         ||||||||||||   |||||||||||||||||||||||||   |||   |
     367 pSerAlaValGluAsnLeuLeuLeuLeuSerLysAlaLysSerValSerS 384

    1145 CGGAGCGCGAGGCGTCCCCGAGCAACAGCTGCCAAGACTCCACGGACACC 1194
         ||||||||:::|||||||||||||||||||||||||||||||||||||||
     384 erGluArgAspAlaSerProSerAsnSerCysGlnAspSerThrAspThr 400

    1195 GAGAGCAACAACGAGGAGCAGCGCAGCGGTCTTATCTACCTGACCAACCA 1244
         ||||||||||||||||||   ||||||||||||||||||||||||||||
     401 GluSerAsnAsnGluGlu...ArgSerGlyLeuIleTyrLeuThrAsnHi 416

    1245 CATCGCCCGACGCGCGCAACGC...GTGTCGCTCAAGGAGGAGCACCGCG 1291
         ||||:::   :::|||::::::   :::|||:::|||||||||   |||
     416 sIleGlyProHisAlaArgAsnGlyIleSerValLysGluGluSerArgG 433

    1292 CCTACGACCTGCTGCGCGCCGCCTCCGAGAACTCGCAGGACGCGCTCCGC 1341
           :::|||:::|||||||||:::::::::|||||||||||||||:::::
     433 lnPheAspValLeuArgAlaGlyThrAspAsnSerGlnAspAlaPheLys 449

    1342 GTGGTCAGCACCAGCGGGGAGCAGATGAAGGTGTACAAGTGCGAACACTG 1391
         |||:::|||:::::::|||||||||:::::::|||||||||||||||||
     450 ValIleSerSerAsnGlyGluGlnValArgValTyrLysCysGluHisCy 466

    1392 CCGGGTGCTCTTCCTGGATCACGTCATGTACACCATCCACATGGGCTGCC 1441
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     466 sArgValLeuPheLeuAspHisValMetTyrThrIleHisMetGlyCysH 483

    1442 ACGGCTTCCGTGATCCTTTTGAGTGCAACATGTGCGGCTACCACAGCCAG 1491
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     483 isGlyPheArgAspProPheGluCysAsnMetCysGlyTyrHisSerGln 499

    1492 GACCGGTACGAGTTCTCGTCGCACATAACGCGAGGGGAGCACCGCTTCCA 1541
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     500 AspArgTyrGluPheSerSerHisIleThrArgGlyGluHisArgPheHi 516

    1542 CATGAGC 1548
         |||||||
```

```
    516 sMetSer 518

seq_name: SwissProt_40:IKAR_ONCMY

seq_documentation_block:
ID   IKAR_ONCMY     STANDARD;      PRT;   522 AA.
AC   O13089; O13088; O13086; O13087;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA-binding protein Ikaros.
GN   IKAROS.
OS   Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC   Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX   NCBI_TaxID=8022;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SHASTA; TISSUE=Thymocytes;
RX   MEDLINE=98056818; PubMed=9394836;
RA   Hansen J.D., Strassburger P., du Pasquier L.;
RT   "Conservation of a master hematopoietic switch gene during vertebrate
RT   evolution: isolation and characterization of Ikaros from teleost and
RT   amphibian species.";
RL   Eur. J. Immunol. 27:3049-3058(1997).
CC   -!- FUNCTION: BINDS AND ACTIVATES THE ENHANCER (DELTA-A ELEMENT) OF
CC       THE CD3-DELTA GENE. FUNCTIONS IN THE SPECIFICATION AND THE
CC       MATURATION OF THE T LYMPHOCYTE. ALSO INTERACTS WITH A CRITICAL
CC       CONTROL ELEMENT IN THE TDT (TERMINAL DEOXYNUCLEOTIDYLTRANSFERASE)
CC       PROMOTER AS WELL AS WITH THE PROMOTERS FOR OTHER GENES
CC       EXPRESSED DURING EARLY STAGES OF B AND T CELL DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS: 8 ISOFORMS; IK-1 (SHOWN HERE), IK-2, IK-3,
CC       IK-4, IK-5, IK-6, IK-7 AND IK-8; ARE PRODUCED BY ALTERNATIVE
CC       SPLICING.
CC   -!- TISSUE SPECIFICITY: EXPRESSION MAINLY LIMITED TO THYMUS, SPLEEN,
CC       AND PRONEPHROS. VERY LOW EXPRESSION IN LIVER. NO EXPRESSION IN
CC       TESTIS, BRAIN, EYE AND MUSCLE.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSION BEGINS AT DAY 3-4 IN THE YOLK SAC
CC       AND AT DAY 5-6 IN THE EMBRYO PROPER.
CC   -!- SIMILARITY: BELONGS TO THE IKAROS FAMILY OF ZINC-FINGER PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U92201; AAB97474.1; -.
DR   EMBL; U92200; AAB53434.1; -.
DR   EMBL; U92198; AAB53432.1; -.
DR   EMBL; U92199; AAB53433.1; -.
DR   HSSP; P08047; 1SP2.
DR   TRANSFAC; T02703; -.
DR   InterPro; IPR000822; Znf-C2H2.
DR   Pfam; PF00096; zf-C2H2; 6.
DR   PRINTS; PR00048; ZINCFINGER.
DR   SMART; SM00355; ZnF_C2H2; 6.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW   DNA-binding; Nuclear protein; Repeat; Alternative splicing.
FT   DOMAIN      125    232       ZINC-FINGERS I.
FT   ZN_FING     125    147       C2H2-TYPE.
FT   ZN_FING     153    175       C2H2-TYPE.
FT   ZN_FING     181    203       C2H2-TYPE.
FT   ZN_FING     209    232       C2H2-TYPE.
FT   DOMAIN      468    520       ZINC-FINGERS II.
FT   ZN_FING     468    490       C2H2-TYPE.
FT   ZN_FING     496    520       C2H2-TYPE.
FT   VARSPLIC     55    148       MISSING (IN ISOFORM IK-2, ISOFORM IK-4
FT                                AND ISOFORM IK-8).
FT   VARSPLIC     55    291       MISSING (IN ISOFORM IK-6).
FT   VARSPLIC    149    291       MISSING (IN ISOFORM IK-5).
FT   VARSPLIC    206    246       MISSING (IN ISOFORM IK-4).
FT   VARSPLIC    206    291       MISSING (IN ISOFORM IK-3).
FT   VARSPLIC    247    291       MISSING (IN ISOFORM IK-7 AND ISOFORM IK-
FT                                8).
SQ   SEQUENCE   522 AA;  57657 MW;  EE9DA9B2A0ECF88A CRC64;


alignment_scores:
            Quality: 1791.50              Length:    531
              Ratio:  3.937                Gaps:     10
 Percent Similarity:  85.687   Percent Identity: 64.407

alignment_block:
 US-08-711-417C-165 x IKAR_ONCMY    ..

 Align seg 1/1  to: IKAR_ONCMY  from: 1  to: 522

       1 ATGGATGCTGACGAGGGTCAAGACATGTCTTTCTCATCAGGGAAGGAAAG 50
         |||:::   :::|||:::|||:::|||||||          |||:::::::||
       1 MetGluMetGluGluAlaGlnGluMetSerGlnMetProGlyArgAspSe 17

      51 CCCCCCTGTAAGCGATACTCCAGATGAGGGCGATGAGCCCATGCCGATCC 100
         |||||||   :::|||:::   :::|||:::|||||||   ||||||||||
      17 rProProProAsnAspValSerGluGluAsnAspGluAlaMetProIleP 34

     101 CCGAGGACCTCTCCACCACCTCGGGAGGACAGCAAAGCTCCAAGAGTGAC 150
         ||||||||||||||:::::::|||:::   |||:::::::::::::::|||
      34 roGluAspLeuSerAlaSerSerAsnLeuGlnHisAsnAsnArgGlyAsp 50

     151 AGAGTC...GTGGCCAGTAATGTTAAAGTAGAGACTCAGAGTGATGAAGA 197
         :::      :::|||   |||:::|||||||||:::::::   ||||||||
      51 LysGluGlyLeuAlaCysAsnIleLysValGluAlaArgCysAspGluGl 67

     198 GAATGGGCGTGCCTGTGAA......ATGAATGGG.........GAAGAAT 232
         |||||||   |||   :::       |||||||||         |||||||
      67 uAsnGlyLeuAlaIleAspMetMetMetAsnGlyGluGluGluGluGluC 84

     233 GTGCGGAGGATTTACGAATGCTTGATGCCTCGGGAGAGAAAATGAATGGC 282
         |||||||||||||||||:::|||||||||||||||   |||:::||||||
      84 ysAlaGluAspLeuArgValLeuAspAlaSerGlyAlaLysValAsnGly 100

     283 TCCCACAGG......GACCAAGGCAGCTCGGCTTTGTCGGGAGTTGGAGG 326
         ||||||            ::::::::::::      |||:::::|||||
     101 SerHisAlaGlyGlyProAspSerLysGlyProTyrSerSerAlaGlyGl 117

     327 CATTCGACTTCCTAACGGAAAACTAAAGTGTGATATCTGTGGGATCATTT 376
         ||||||||||||||||||||||||||||||||||||||||||||||||:::|
     117 yIleArgLeuProAsnGlyLysLeuLysCysAspIleCysGlyIleValC 134

     377 GCATCGGGCCCAATGTGCTCATGGTTCACAAAAGAAGCCACACTGGAGAA 426
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 ysIleGlyProAsnValLeuMetValHisLysArgSerHisThrGlyGlu 150

     427 CGGCCCTTCCAGTGCAATCAGTGCGGGGCCTCATTCACCCAGAAGGGCAA 476
         ||||||||||||||:::|||||||||||||||||||||||||||||||||
     151 ArgProPheGlnCysThrGlnCysGlyAlaSerPheThrGlnLysGlyAs 167

     477 CCTGCTCCGGCACATCAAGCTGCATTCCGGGGAGAAGCCCTTCAAATGCC 526
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     167 nLeuLeuArgHisIleLysLeuHisSerGlyGluLysProPheLysCysH 184

     527 ACCTCTGCAACTACGCCTGCCGCCGGAGGGACGCCCTCACTGGCCACCTG 576
         ||||||||||||||||||||||||||||||||||:::|||||||||
     184 isLeuCysAsnTyrAlaCysArgArgArgAspAlaLeuSerGlyHisLeu 200

     577 AGGACGCACTCCGTTGGTAAACCTCACAAATGTGGATATTGTGGCCGAAG 626
         ||||||||||||||||||||||||||||||||:::|||||||||||||
     201 ArgThrHisSerValGlyLysProHisLysCysAlaTyrCysGlyArgSe 217
```

```
 627 CTATAAACAGCGAACGTCTTTAGAGGAACATAAAGAGCGCTGCCACAACT 676
     ||||||||||||||:::|||||||||||||||||||||||||||||||||||
 217 rTyrLysGlnArgSerSerLeuGluGluHisLysGluArgCysHisAsnT 234

 677 ACTTGGAAAGCATGGGCCTTCCGGGCACACTGTACCCAGTCATTAAAGAA 726
     |||||:::   |||||||||   ::::::::::|||   |||:::||||||
 234 yrLeuGlnCysMetGlyLeuGlnAsnSerIleTyrThrValValLysGlu 250

 727 GAAACTAAGCACAGTGAAATGGCAGAAGACCTGTGCAAGATAGGATCAGA 776
     |||:::::::::::::|||:::   |||||||||   ::::::||||||::
 251 GluSerAsnGlnAsnGluGlnArgGluAspLeuSerGlnMetGlySerLy 267

 777 GAGATCTCTCGTGCTGGACAGACTAGCAAGTAATGTCGCCAAACGTAAGA 826
     :|||:::|||||||||||||||||||:::|||||||||||||||||||||
 267 sArgAlaLeuValLeuAspArgLeuAlaAsnAsnValAlaLysArgLysS 284

 827 GCTCTATGCCTCAGAAATTTCTTGGGGACAAGGGCCTGTCCGACACGCCC 876
     ||:::|||||||||||||||:::|||:::|||   :::|||:::
 284 erThrMetProGlnLysPheValGlyGluLysArgPheSerAsnIleSer 300

 877 TACGACAGTGCCACGTACGAGAAGGAGAACGAAATGATGAAGTCCCACGT 926
     ::::::::::::              :::|||:::|||:::   |||||
 301 PheGluGlyGly.............ProGlyGluLeuMetGlnProHisVa 313

 927 GATGGACCAAGCCATCAACAACGCCATCAACTACCTGGGGGCCGAGTCCC 976
     |:::|||||||||||||||:::|||||||||||||||||||||||||||||
 313 lIleAspGlnAlaIleAsnSerAlaIleAsnTyrLeuGlyAlaGluSerL 330

 977 TGCGCCCGCTGGTGCAGACGCCCCCGGGCGGTTCCGAGGTGGTCCCGGTC 1026
     ||||||||||:::||||||   |||   :::|||:::   :::   |||
 330 euArgProLeuIleGlnThrSerProThrSerSerAsp...MetGlyVal 345

1027 ATCAGCCCGATGTACCAGCTGCACAGGCGCTCGGAGGGCACCCCGCGCTC 1076
     ::::::   ||||||   |||||||::              |||   ::
 346 MetGlySerMetTyrProLeuHisLys.............ProProAlaGl 358

1077 CAACCAC......TCGGCCCAGGACAGCGCCGTGGAGTACCTGCTGCTGC 1120
     ::::|||      ||||||:::|||||||||:::|||   ||||||||||
 358 uGlyHisGlyLeuSerAlaLysAspSerAlaAlaGluAsnLeuLeuLeuL 375

1121 TCTCCAAGGCCAAGTTGGTGCCCTCGGAGCGCGAGGCGTCCCCGAGCAAC 1170
     ||:::|||:::|||   :::   ||||||:::::::::|||||||||:::
 375 euAlaLysSerLysSerAlaSerSerGluLysAspGlySerProSerHis 391

1171 AGCTGCCAAGACTCCACGGACACCGAGAGCAACAACGAGGAGCAG..... 1215
     |||   ||||||||||||||||||||||||||||||||||||:::
 392 SerGlyGlnAspSerThrAspThrGluSerAsnAsnGluGluLysAlaGl 408

1216 .......CGCAGCGGTCTTATCTACCTGACCAACCACATCGCCCGACGCG 1258
             ||||||||||||||||||||||||||||||:::      :
 408 yValGlyAlaSerGlyLeuIleTyrLeuThrAsnHisIleThrSerGlyV 425

1259 CGCAACGCGTGTCGCTC............AAGGAGGAGCACCGCGCCTAC 1296
     ::::::::      |||           :::|||::::::|||   |||
 425 alArgAsnGlyValLeuProLeuValLysGluGluGlnGlnArgGlnTyr 441

1297 GACCTGCTGCGCGCCGCCTCCGAGAACTCGCAGGACGCGCTCCGCGTGGT 1346
     :::   :::||||||:::   |||   :::::::::::::::::|||::
 442 GluAlaMetArgAlaSerIleGluIleAlaSerGluGlyPheLysValLe 458

1347 CAGCACCAGCGGGGAGCAGATGAAGGTGTACAAGTGCGAACACTGCCGGG 1396
     :|||   :::|||||||:::::::::|||:::|||||||||||||||:
 458 uSerGlyGluGlyGluGlnValArgAlaTyrArgCysGluHisCysArgI 475

1397 TGCTCTTCCTGGATCACGTCATGTACACCATCCACATGGGCTGCCACGGC 1446
     ::|||||||||||||||||||||||||||||||||||||||||||||||||
 475 leLeuPheLeuAspHisValMetTyrThrIleHisMetGlyCysHisGly 491

1447 TTCCGTGATCCTTTTGAGTGCAACATGTGCGGCTACCACAGCCAGGACCG 1496
     |||||||||||||||||||||||:::||||||:::::::||||||||||
 492 PheArgAspProPheGluCysAsnLeuCysGlyHisArgSerGlnAspAr 508

1497 GTACGAGTTCTCGTCGCACATAACGCGAGGGGAGCACCGCTTC 1539
```

```
     ||||||||||||||||||:::||||||||||||||||||:::
 508 gTyrGluPheSerSerHisMetThrArgGlyGluHisArgTyr 522
```

```
seq_name: SwissProt_40:HELI_HUMAN

seq_documentation_block:
ID   HELI_HUMAN     STANDARD;      PRT;   526 AA.
AC   Q9UKS7;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Zinc finger protein Helios.
GN   ZNFN1A2 OR HELIOS.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20009537; PubMed=10541817;
RA   Hosokawa Y., Maeda Y., Seto M.;
RT   "Human Helios, an Ikaros-related zinc finger DNA binding protein: cDNA
RT   cloning and tissue expression pattern.";
RL   Immunogenetics 50:106-108(1999).
CC   -!- FUNCTION: ASSOCIATES WITH IKAROS AT CENTROMERIC HETEROCHROMATIN.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE IKAROS FAMILY OF ZINC-FINGER PROTEINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF130863; AAF09441.1; -.
DR   HSSP; P15822; 1BBO.
DR   MIM; 606234; -.
DR   InterPro; IPR000822; Znf-C2H2.
DR   Pfam; PF00096; zf-C2H2; 5.
DR   SMART; SM00355; ZnF_C2H2; 6.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW   DNA-binding; Nuclear protein; Repeat.
FT   ZN_FING     112    134       C2H2-TYPE.
FT   ZN_FING     140    162       C2H2-TYPE.
FT   ZN_FING     168    190       C2H2-TYPE.
FT   ZN_FING     196    219       C2H2-TYPE.
FT   ZN_FING     471    493       C2H2-TYPE.
FT   ZN_FING     499    523       C2H2-TYPE.
SQ   SEQUENCE   526 AA;  59558 MW;  A834CB0E0E096647 CRC64;


alignment_scores:
            Quality: 1380.00              Length:    519
              Ratio:  3.416                Gaps:     12
 Percent Similarity:  77.842   Percent Identity: 53.757

alignment_block:
 US-08-711-417C-165 x HELI_HUMAN   ..

 Align seg 1/1  to: HELI_HUMAN  from: 1  to: 526

      67 ACTCCAGATGAGGGCGATGAGCCCATGCCGATCCCCGAGGACCTC...TC 113
         :::|||::::::      :::   |||   |||      ||||||   ||
      17 SerProGluArgGluHisSerAsnMetAlaIle......AspLeuThrSe 31

     114 CACCACCTCGGGAGGACAGCAAAGCTCCAAGAGTGACAGAGTCGTGGCCA 163
         |:::|||   :::||||||:::::|||   |||      :::   ::::
      31 rSerThrProAsnGlyGlnHisAlaSerProSerHisMetThrSerThrA 48
```

```
164 GTAATGTTAAAGTAGAGACTCAGAGTGATGAAGAG..............  198
    :::::||||||:::|||   ||||||||||||||||
 48 snSerValLysLeuGluMetGlnSerAspGluGluCysAspArgLysPro  64

199 AATGGGCGTGCCTGTGAAATGAATGGGGAAGAATGTGCGGAGGATTTACG  248
     :::|||      |||:::::::|||::::::   :::::::::|||::
 65 LeuSerArgGluAsnGluIleArgGlyHisAspGluGlySerSerLeuGl  81

249 A......ATGCTTGATGCCTCGGGAGAGAAAATGAATGGCTCCCACAGGG  292
    :     :::::::::::|||            |||   :::   ::::
 81 uGluProLeuIleGluSerSerGluValAlaAspAsnArgLysValGlnG  98

293 ACCAAGGCAGCTCGGCTTTGTCGGGAGTTGGAGGCATTCGACTTCCTAAC  342
    ::              |||:::|||   ||||||||||||||||||||
 98 lu..............LeuGlnGlyGluGlyGlyIleArgLeuProAsn  109

343 GGAAAACTAAAGTGTGATATCTGTGGGATCATTTGCATCGGGCCCAATGT  392
    ||||||||||||||||:::|||||||:::::|||||||||||||||||
110 GlyLysLeuLysCysAspValCysGlyMetValCysIleGlyProAsnVa  126

393 GCTCATGGTTCACAAAAGAAGCCACACTGGAGAACGGCCCTTCCAGTGCA  442
    ||||||||||||||||||||||||||||||||||||||||:::||||
126 lLeuMetValHisLysArgSerHisThrGlyGluArgProPheHisCysA  143

443 ATCAGTGCGGGGGCCTCATTCACCCAGAAGGGCAACCTGCTCCGGCACATC  492
    ||||||||||||||||||||||||||||||||||||||||||||||||||
143 snGlnCysGlyAlaSerPheThrGlnLysGlyAsnLeuLeuArgHisIle  159

493 AAGCTGCATTCCGGGGAGAAGCCCTTCAAATGCCACCTCTGCAACTACGC  542
    ||||||||||||||||||||||||||||||   :::|||:::|||||
160 LysLeuHisSerGlyGluLysProPheLysCysProPheCysSerTyrAl  176

543 CTGCCGCCGGAGGGACGCCCTCACTGGCCACCTGAGGACGCACTCCGTTG  592
    ||||||||||||||||||||||||||||||||||||||||||||||||
176 aCysArgArgArgAspAlaLeuThrGlyHisLeuArgThrHisSerValG  193

593 GTAAACCTCACAAATGTGGATATTGTGGCCGAAGCTATAAACAGCGAACG  642
    ||||||||||||||||:::|||||||||||||||||||||||||||:::
193 lyLysProHisLysCysAsnTyrCysGlyArgSerTyrLysGlnArgSer  209

643 TCTTTAGAGGAACATAAAGAGCGCTGCCACAACTACTTGGAAAGCATGGG  692
    ||||||||||||||||||||||||||||||||||||||:::::::::::
210 SerLeuGluGluHisLysGluArgCysHisAsnTyrLeuGlnAsnValSe  226

693 CCTTCCGGGCACA.....................CTGTACCCAGTCATTA  721
    ::::   ::::::                    :::   |||:::
226 rMetGluAlaAlaGlyGlnValMetSerHisHisValProProMet....  241

722 AAGAAGAAACTAAGCACAGTGAA......ATGGCAGAAGACCTGTGCAAG  765
      |||:::   |||::::::|||      |||   :::::::::
242 ..GluAspCysLysGluGlnGluProIleMetAspAsnAsnIleSerLeu  257

766 ATAGGATCAGAGAGATCTCTCGTGCTGGACAGACTAGCAAGTAATGTCGC  815
    :::      ||||||      |||:::::::::|||:::::|||:::::
258 ValProPheGluArgProAlaValIleGluLysLeuThrGlyAsnMetGl  274

816 CAAACGTAAGAGCTCTATGCCTCAGAAATTTCTTGGGGACAAG.......  858
    :|||||||||||||||   ||||||||||||:::|||:::|||
274 yLysArgLysSerSerThrProGlnLysPheValGlyGluLysLeuMetA  291

859 .....GGCCTGTCCGACACGCCCTACGAC...AGTGCCACGTACGAGAAG  900
         :::     |||      :::|||   :::   ||||||||||||
291 rgPheSerTyrProAspIleHisPheAspMetAsnLeuThrTyrGluLys  307

901 GAGAACGAAATGATGAAGTCCCACGTGATGGACCAAGCCATCAACAACGC  950
    |||   |||:::|||:::|||||||:::|||||||||||||||||||||
308 GluAlaGluLeuMetGlnSerHisMetMetAspGlnAlaIleAsnAsnAl  324

951 CATCAACTACCTGGGGGCCGAGTCCCTGCGCCCGCTGGTGCAGACGCCCC 1000
    ||||:::|||||||||||||||:::|||:::|||||||:::|||   ||||
324 aIleThrTyrLeuGlyAlaGluAlaLeuHisProLeuMetGlnHisProP  341

1001 CGGGCGGT...TCCGAGGTGGTCCCGGTCATC............AGCCCG 1035
```

```
     ||:::      :::|||||||:::|||||||||           |||
 341 roSerThrIleAlaGluValAlaProValIleSerSerAlaTyrSerGln  357

1036 ATGTACCAGCTGCACAGGCGCTCGGAGGGCACCCCGCGCTCCAACCACTC 1085
     :::|||:::   :::|||   ::::::         |||::::::   ::
 358 ValTyrHisProAsnArgIleGluArgProIleSerArgGluThrAlaAs  374

1086 GGCCCAGGACAGCGCCGTGGAGTACCTGCTGCTGCTCTCCAAGGCCAAGT 1135
     :::::::::::::   ::::::      :::   |||   :::   |||
 374 pSerHisGluAsnAsnMetAspGlyProIleSerLeuIleArgProLysS  391

1136 TGGTGCCCTCGGAGCGCGAGGCGTCCCCGAGCAACAGCTGCCAAGACTCC 1185
          |||:::|||||||||||||||||||||||||||||||   ||||||
 391 erArgProGlnGluArgGluAlaSerProSerAsnSerCysLeuAspSer  407

1186 ACGGACACCGAGAGCAACAACGAGGAGCAGCGCAGCGGTCTTATCTACCT 1235
     ||||||:::|||||||:::::::::::::::::::|||      :::
 408 ThrAspSerGluSerSerHisAspAspHisGlnSerTyrGlnGlyHisPr  424

1236 GACCAACCACATCGCCCGACGCGCGCAACGCGTGTCGCTCAAGGAGGAGC 1285
      :::   :::      |||:::   :::   :::   :::||||||:::
 424 oAlaLeuAsnProLysArgLysGlnSerProAlaTyrMetLysGluAspV  441

1286 ACCGCGCCTACGACCTGCTGCGCGCCGCCTCCGAGAACTCGCAGGACGCG 1335
       :::|||   |||      :::|||   :::   :::   :::||||
 441 alLysAlaLeuAspThrThrLysAlaProLysGlySerLeuLysAspIle  457

1336 CTCCGCGTGGTCAGCACCAGCGGGGAGCAGATGAAGGTGTACAAGTGCGA 1385
       :::|||   :::   :::|||||||||:::::::::::|||||||||
 458 TyrLysValPheAsnGlyGluGlyGluGlnIleArgAlaPheLysCysGl  474

1386 ACACTGCCGGGTGCTCTTCCTGGATCACGTCATGTACACCATCCACATGG 1435
     ||||||||||||||||||||||||||||||||||||||||||||||||
 474 uHisCysArgValLeuPheLeuAspHisValMetTyrThrIleHisMetG  491

1436 GCTGCCACGGCTTCCGTGATCCTTTTGAGTGCAACATGTGCGGCTACCAC 1485
     ||||||||||:::|||||||||:::|||||||||:::|||||||||:::
 491 lyCysHisGlyTyrArgAspProLeuGluCysAsnIleCysGlyTyrArg  507

1486 AGCCAGGACCGGTACGAGTTCTCGTCGCACATAACGCGAGGGGAGCACCG 1535
     ||||||||||||||||||||||||||||||||:::|||||||||||||
 508 SerGlnAspArgTyrGluPheSerSerHisIleValArgGlyGluHisTh  524

1536 CTTCCAC 1542
      ||||||
 524 rPheHis 526
```

```
seq_name: SwissProt_40:HELI_MOUSE

seq_documentation_block:
ID   HELI_MOUSE     STANDARD;      PRT;   526 AA.
AC   P81183;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Zinc finger protein Helios.
GN   ZNFN1A2 OR HELIOS.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 246-267 AND 289-306.
RC   STRAIN=BALB/C; TISSUE=Thymus;
RX   MEDLINE=98180981; PubMed=9512513;
RA   Hahm K., Cobb B.S., McCarty A.S., Brown K.E., Klug C.A., Lee R.,
RA   Akashi K., Weissman I.L., Fisher A.G., Smale S.T.;
RT   "Helios, a T cell-restricted Ikaros family member that quantitatively
RT   associates with Ikaros at centromeric heterochromatin.";
RL   Genes Dev. 12:782-796(1998).
CC   -!- FUNCTION: ASSOCIATES WITH IKAROS AT CENTROMERIC HETEROCHROMATIN.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A AND B (SHOWN HERE); ARE
```

```
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: RESTRICTED TO THE T-CELL LINEAGE. ABUNDANT IN
CC       THYMUS, LOW EXPRESSION IN BONE MARROW AND BRAIN AND
CC       NO DETECTABLE EXPRESSION IN SPLEEN, LIVER, KIDNEY OR MUSCLE.
CC   -!- SIMILARITY: BELONGS TO THE IKAROS FAMILY OF ZINC-FINGER PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF044257; AAC00513.1; -.
DR   HSSP; P15822; 1BBO.
DR   MGD; MGI:1342541; Znfn1a2.
DR   InterPro; IPR000822; Znf-C2H2.
DR   Pfam; PF00096; zf-C2H2; 5.
DR   PRINTS; PR00048; ZINCFINGER.
DR   SMART; SM00355; ZnF_C2H2; 6.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW   Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW   DNA-binding; Nuclear protein; Repeat; Alternative splicing.
FT   DOMAIN      112    219       ZINC-FINGERS I.
FT   ZN_FING     112    134       C2H2-TYPE.
FT   ZN_FING     140    162       C2H2-TYPE.
FT   ZN_FING     168    190       C2H2-TYPE.
FT   ZN_FING     196    219       C2H2-TYPE.
FT   DOMAIN      471    523       ZINC-FINGERS II.
FT   ZN_FING     471    493       C2H2-TYPE.
FT   ZN_FING     499    523       C2H2-TYPE.
FT   VARSPLIC    111    136       MISSING (IN ISOFORM A).
SQ   SEQUENCE   526 AA;  59388 MW;  EFD827FE8916108E CRC64;


alignment_scores:
            Quality: 1374.50              Length:    518
              Ratio:  3.454                Gaps:     12
 Percent Similarity:  76.834   Percent Identity: 53.668

alignment_block:
 US-08-711-417C-165 x HELI_MOUSE   ..

 Align seg 1/1  to: HELI_MOUSE  from: 1  to: 526

      76 GAGGGCGATGAGCCCATGCCGATCCCCGAGGACCTCTCCACCACCTCGGG 125
         ||||||::::::               :::      |||:::|||   ::
      19 GluGlyGluHisAlaAsnMetAlaIleAspLeuThrSerSerThrProAs 35

     126 AGGACAGCAAAGCTCCAAGAGTGACAGAGTCGTGGCCAGTAATGTTAAAG 175
         :||||||::::::|||   |||      :::   :::::::::||||||:
      35 nGlyGlnHisAlaSerProSerHisMetThrSerThrAsnSerValLysL 52

     176 TAGAGACTCAGAGTGATGAAGAG..............AATGGGCGTGCC 210
         ::|||   ||||||||||||||                ::::|||
      52 euGluMetGlnSerAspGluGluCysAspArgGlnProLeuSerArgGlu 68

     211 TGTGAAATGAATGGGGAAGAATGTGCGGAGGATTTACGA......ATGCT 254
            |||::::::|||::::::   :::::::::|||:::      :::::
      69 AspGluIleArgGlyHisAspGluGlySerSerLeuGluGluProLeuIl 85

     255 TGATGCCTCGGGAGAGAAAATGAATGGCTCCCACAGGGACCAAGGCAGCT 304
         :::::::|||            |||   :::   :::|||
      85 eGluSerSerGluValAlaAspAsnArgLysValGlnAsp..........  98

     305 CGGCTTTGTCGGGAGTTGGAGGCATTCGACTTCCTAACGGAAAACTAAAG 354
               |||:::|||   ||||||||||||||||||||||||||||||||
      99 .....LeuGlnGlyGluGlyGlyIleArgLeuProAsnGlyLysLeuLys 113

     355 TGTGATATCTGTGGGATCATTTGCATCGGGCCCAATGTGCTCATGGTTCA 404
         ||||||:::||||||::::::||||||||||||||||||||||||||||||

     114 CysAspValCysGlyMetValCysIleGlyProAsnValLeuMetValHi 130

     405 CAAAAGAAGCCACACTGGAGAACGGCCCTTCCAGTGCAATCAGTGCGGGG 454
         ||||||||||||||||||||||||||||||:::||||||||||||||||
     130 sLysArgSerHisThrGlyGluArgProPheHisCysAsnGlnCysGlyA 147

     455 CCTCATTCACCCAGAAGGGCAACCTGCTCCGGCACATCAAGCTGCATTCC 504
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     147 laSerPheThrGlnLysGlyAsnLeuLeuArgHisIleLysLeuHisSer 163

     505 GGGGAGAAGCCCTTCAAATGCCACCTCTGCAACTACGCCTGCCGCCGGAG 554
         ||||||||||||||||||||   :::|||:::||||||||||||||||||
     164 GlyGluLysProPheLysCysProPheCysSerTyrAlaCysArgArgAr 180

     555 GGACGCCCTCACTGGCCACCTGAGGACGCACTCCGTTGGTAAACCTCACA 604
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     180 gAspAlaLeuThrGlyHisLeuArgThrHisSerValGlyLysProHisL 197

     605 AATGTGGATATTGTGGCCGAAGCTATAAACAGCGAACGTCTTTAGAGGAA 654
         |||||:::||||||||||||||||||||||||||:::|||||||||||||
     197 ysCysAsnTyrCysGlyArgSerTyrLysGlnArgSerSerLeuGluGlu 213

     655 CATAAAGAGCGCTGCCACAACTACTTGGAAAGCATGGGCCTTCCGGGCAC 704
         ||||||||||||||||||||||||||:::::::::::::::   :::::
     214 HisLysGluArgCysHisAsnTyrLeuGlnAsnValSerMetGluAlaAl 230

     705 A.....................CTGTACCCAGTCATTAAAGAAGAAACTA 733
         :                     :::   |||:::      |||:::   |
     230 aGlyGlnValMetSerHisHisValProProMet......GluAspCysL 245

     734 AGCACAGTGAA......ATGGCAGAAGACCTGTGCAAGATAGGATCAGAG 777
         ||:::::::|||      |||   :::::::::      :::      |||
     245 ysGluGlnGluProIleMetAspAsnAsnIleSerLeuValProPheGlu 261

     778 AGATCTCTCGTGCTGGACAGACTAGCAAGTAATGTCGCCAAACGTAAGAG 827
         |||      |||:::::::::|||:::::::|||::::::|||||||||||
     262 ArgProAlaValIleGluLysLeuThrAlaAsnMetGlyLysArgLysSe 278

     828 CTCTATGCCTCAGAAATTTCTTGGGGACAAG............GGCCTGT 865
         ||||   ||||||||||||:::|||:::|||            :::
     278 rSerThrProGlnLysPheValGlyGluLysLeuMetArgPheSerTyrP 295

     866 CCGACACGCCCTACGAC...AGTGCCACGTACGAGAAGGAGAACGAAATG 912
           |||      :::|||   :::   |||||||||||||||   |||:::
     295 roAspIleHisPheAspMetAsnLeuThrTyrGluLysGluAlaGluLeu 311

     913 ATGAAGTCCCACGTGATGGACCAAGCCATCAACAACGCCATCAACTACCT 962
         |||:::|||||||:::|||||||||||||||||||||||||||||:::|||||
     312 MetGlnSerHisMetMetAspGlnAlaIleAsnAsnAlaIleThrTyrLe 328

     963 GGGGGCCGAGTCCCTGCGCCCGCTGGTGCAGACGCCCCCGGGCGGT...T 1009
         ||||||||||:::|||:::|||||||:::|||      |||:::      :
     328 uGlyAlaGluAlaLeuHisProLeuMetGlnHisAlaProSerThrIleA 345

    1010 CCGAGGTGGTCCCGGTCATC............AGCCCGATGTACCAGCTG 1047
         ::||||||:::|||||||||            |||   :::|||:::
     345 laGluValAlaProValIleSerSerAlaTyrSerGlnValTyrHisPro 361

    1048 CACAGGCGCTCGGAGGGCACCCCGCGCTCCAACCACTCGGCCCAGGACAG 1097
         :::|||   ::::::        |||::::::   ::::::::::::::
     362 AsnArgIleGluArgProIleSerArgGluThrSerAspSerHisGluAs 378

    1098 CGCCGTGGAGTACCTGCTGCTGCTCTCCAAGGCCAAGTTGGTGCCCTCGG 1147
         :   ::::::            |||   :::   |||      |||:::|
     378 nAsnMetAspGlyProThrSerLeuIleArgProLysSerArgProGlnG 395

    1148 AGCGCGAGGCGTCCCCGAGCAACAGCTGCCAAGACTCCACGGACACCGAG 1197
         ||||||||||||||||||||||||||||   ||||||||||||:::|||
     395 luArgGluAlaSerProSerAsnSerCysLeuAspSerThrAspSerGlu 411

    1198 AGCAACAACGAGGAGCAGCGCAGCGGTCTTATCTACCTGACCAAC..... 1242
         |||:::::::::::::::::|||         |||      |||
     412 SerSerHisAspAspArgGlnSer.........TyrGlnGlyAsnProAl 425
```

```
    1243 ....CACATCGCCCGACGCGCGCAACGCGTGTCGCTCAAGGAGGAGCACC 1288
           ::      |||:::   :::   :::   :::||||||:::   :
     425 aLeuAsnProLysArgLysGlnSerProAlaTyrMetLysGluAspValL 442

    1289 GCGCCTACGACCTGCTGCGCGCCGCCTCCGAGAACTCGCAGGACGCGCTC 1338
         ::|||   |||      :::|||   :::   :::   :::|||
     442 ysAlaLeuAspAlaThrLysAlaProLysGlySerLeuLysAspIleTyr 458

    1339 CGCGTGGTCAGCACCAGCGGGGAGCAGATGAAGGTGTACAAGTGCGAACA 1388
         :::|||   :::   :::||||||||||:::::::::::||||||||||
     459 LysValPheAsnGlyGluGlyGluGlnIleArgAlaPheLysCysGluHi 475

    1389 CTGCCGGGTGCTCTTCCTGGATCACGTCATGTACACCATCCACATGGGCT 1438
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     475 sCysArgValLeuPheLeuAspHisValMetTyrThrIleHisMetGlyC 492

    1439 GCCACGGCTTCCGTGATCCTTTTGAGTGCAACATGTGCGGCTACCACAGC 1488
         ||||||||:::||||||||||:::||||||||||:::||||||||||:::|||
     492 ysHisGlyTyrArgAspProLeuGluCysAsnIleCysGlyTyrArgSer 508

    1489 CAGGACCGGTACGAGTTCTCGTCGCACATAACGCGAGGGGAGCACCGCTT 1538
         ||||||||||||||||||||||||||||||:::||||||||||||   ||
     509 GlnAspArgTyrGluPheSerSerHisIleValArgGlyGluHisThrPh 525

    1539 CCAC 1542
         ||||
     525 eHis 526
```

seq_name: SwissProt_40:HUNB_CLOAL

seq_documentation_block:

```
ID   HUNB_CLOAL     STANDARD;      PRT;   485 AA.
AC   O96785;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hunchback protein (Fragment).
GN   HB.
OS   Clogmia albipunctata (Mothmidge).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Psychodoidea;
OC   Psychodidae; Clogmia.
OX   NCBI_TaxID=85120;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rohr K.B., Tautz D., Sander K.;
RT   "Segmentation gene expression in the mothmidge Clogmia albipunctata
RT   (Diptera, Psychodidae) and other primitive dipterans.";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: GAP CLASS SEGMENTATION PROTEIN THAT CONTROLS DEVELOPMENT
CC       OF HEAD STRUCTURES (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: BELONGS TO THE HUNCHBACK FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ131041; CAA10281.1; -.
DR   InterPro; IPR000822; Znf-C2H2.
DR   Pfam; PF00096; zf-C2H2; 6.
DR   SMART; SM00355; ZnF_C2H2; 6.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.
KW   Developmental protein; Gap protein; Zinc-finger;
KW   Metal-binding; DNA-binding; Repeat; Nuclear protein.
FT   NON_TER       1      1
FT   DOMAIN        87     196       ZINC FINGERS I.
FT   DOMAIN        432    484       ZINC FINGERS II.
FT   ZN_FING       87     109       C2H2-TYPE.
FT   ZN_FING       116    138       C2H2-TYPE.
FT   ZN_FING       144    166       C2H2-TYPE.
FT   ZN_FING       172    196       C2H2-TYPE.
FT   ZN_FING       432    454       C2H2-TYPE.
FT   ZN_FING       460    484       C2H2-TYPE.
SQ   SEQUENCE   485 AA;  55367 MW;  8BEAC3B3C3B0C37C CRC64;
```

alignment_scores:

```
            Quality:  334.50              Length:    577
              Ratio:  1.216                 Gaps:     25
 Percent Similarity:  47.660   Percent Identity: 22.530
```

alignment_block:

US-08-711-417C-165 x HUNB_CLOAL ..

Align seg 1/1 to: HUNB_CLOAL from: 1 to: 485

```
      34 TCATCAGGGAAGGAAAGCCCCCCTGTAAGCGATACTCCAGATGAGGGCGA 83
         |||:::::::::::::::|||      ::::::      ::::::   ::
       3 SerThrAlaArgLysThrProGluLysAspSerLeuLysGlnAspGlnAs 19

      84 TGAGCCCATGCCGATCCCCGAGGACCTCTCCACCACCTCGGGAGGACAGC 133
         ::::   :::      |||                |||:::|||:::||||
      19 nGlnLeuLeuLysThrPro.........IleGlnThrAsnGlyAsnGlnG 33

     134 AAAGCTCCAAGAGT..................................... 147
         ||||||:::   :::
      33 lnSerThrPheAspSerGlyGluAspSerHisSerMetProAspSerAsp 49

     148 ......GACAGAGTCGTGGCCAGTAATGTTAAAGTAGAGACTCAGAGTGA 191
               :::   |||:::::::::::::::   |||:::   ::::::||
      50 LeuLeuGluProValIleThrAspGlyAlaAspValAspAspGluAsnAs 66

     192 TGAAGAGAATGGGCGTGCCTGTGAAATGAATGGGGAAGAATGTGCGGAGG 241
         |   |||:::                                    :::|
      66 pAlaGluGluAsp..............................AspA 72

     242 ATTTACGAATGCTTGATGCCTCGGGAGAGAAAATGAATGGCTCCCACAGG 291
         ||:::|||                        |||:::|||   ||||||
      72 spIleArgThr..............ProLysIleAsn...SerHis... 81

     292 GACCAAGGCAGCTCGGCTTTGTCGGGAGTTGGAGGCATTCGACTTCCTAA 341

      81 ..................................................  81

     342 CGGAAAACTA.........AAGTGTGATATCTGTGGGATCATTTGCATCG 382
          ||||||:::         ||||||      |||   :::|||::::::
      82 .GlyLysMetLysThrTyrLysCysLysGlnCysAspPheIleAlaValT 98

     383 GGCCCAATGTGCTCATGGTTCACAAAAGAAGCCAC...ACTGGAGAACGG 429
                   :::      |||:::|||   |||         |||:::
      98 hrLysLeuSerPheTrpGluHisAsnArgIleHisIleLysProGluLys 114

     430 CCCTTCCAGTGCAATCAGTGCGGGGCCTCATTCACCCAGAAGGGCAACCT 479
           ::::::|||:::::::|||                  |||   :::||
     115 MetLeuLysCysGlnLysCysProPheIleThrGluTyrLysHisHisLe 131

     480 GCTCCGGCACATCAAGCTGCATTCCGGGGAGAAGCCCTTCAAATGCCACC 529
         |      |||::::::   |||:::|||:::||||||||||:::|||
     131 uGluTyrHisLeuArgAsnHisAsnGlySerLysProPheGlnCysLysG 148

     530 TCTGCAACTACGCCTGCCGCCGGAGGGACGCCCTCACTGGCCACCTGAGG 579
           |||||||||:::|||   :::::::::   |||:::::::|||::::::
     148 lnCysAsnTyrSerCysValAsnLysSerMetLeuAsnSerHisMetLys 164

     580 ACGCACTCCGTTGGTAAACCTCACAAATGTGGATATTGTGGCCGAAGCTA 629
         :::||||||              ::::::|||      |||:::   :::
     165 SerHisSerAsnIleTyrGlnTyrArgCysLysAspCysAsnTyrAlaTh 181
```

```
 630 TAAACAGCGAACGTCTTTAGAGGAACAT...AAAGAGCGCTGCCACAACT 676
 181 rLysTyrCysHisSerLeuLysLeuHisLeuArgLysTyrSerHisAsnP 198

 677 ACTTGGAAAGCATGGGCCTTCCGGGCACACTGTACCCAGTCATTAAAGAA 726
 198 roProMetValLeuAsnTyrAspGlyThrProAsnProLeu..... 211

 727 GAAACTAAGCACAGTGAAATGGCAGAAGACCTGTGCAAGATAGGATCAGA 776
 211 ..... 211

 777 GAGATCTCTCGTGCTGGACAGACTAGCAAGTAATGTCGCCAAACGTAAGA 826
 212 .....ArgIleIleAspValTyrGlyThrArgArgG 222

 827 GCTCTATGCCTCAG...AAATTTCTTGGGGACAAGGGCCTGTCCGACACG 873
 222 ly......ProLysValLysPheHisLysAspGluGly...Gly..HisA 235

 874 CCCTACGACAGTGCCACGTACGAGAAGGAGAACGAAATGATGAAGTCCCA 923
 235 snLeuLeuAsnSerAsnIleAsnThrSerArgArgSerLysSerGlyLys 251

 924 CGTGATGGA...CCAAGCCATCAACAACGCCATCAACTACCTGGGGGCCG 970
 252 ArgAspSerPheProAsnPheGluGlnSerGln..... 262

 971 AGTCCCTGCGCCCGCTGGTGCAGACGCCCCCGGGCGGTTCCGAGGTGGTC 1020
 262 ..... 262

1021 CCGGTCATCAGCCCGATGTACCAGCTGCACAGGCGCTCGGAGGGCACCCC 1070
 263 .....HisValProThrProProSerSerGlnAlaLeuAlaMetLeuPro 277

1071 GCGCTC.....CAACCACTCGGCCCAGG 1093
 278 AsnLeuAlaAsnIlePheGlnGlnSerProSerMetProLeuPheProTy 294

1094 ACAGCGCCGTGGAGTACCTGCTGCTGCTCTCCAAGGCCAAGTTGGTGCCC 1143
 294 rLeuAsn.LeuAsnPheHisHisIleLeuAlaGlnGlnLysAlaAlaLeu 310

1144 TCGGAGCGCGAGGCGTCCCCGAGC.....AACAG 1172
 311 Ser......GlnIleSerProSerIleAsnGlyTrpGlnAsnGluGluAs 325

1173 CTGCCAAGACTCCACGGACACCGAGAGCAACAACGAGGAGCAGCGCAGCG 1222
 325 nCysAsnGluGluGluThrProGluLysGluGluAspProLysArgMetS 342

1223 GTCTTATCTACCTGACCAACCAC......ATCGCCCGACGCGCGCAACGC 1266
 342 erAlaLeuAspLeuSerSerAsnProSerThrProSerThrValSerGln 358

1267 GTGTCGCTCAAGGAGGAGCACCGCGCCTAC......GACCTGCTGCGCGC 1310
 359 ValLysHisLysArgLysGlyArgAlaPheLysLeuGluLeuMetLysGl 375

1311 CGCCTCCGAGAACTCGCAG...GACGCGCTCCGCGTGGTC..... 1347
 375 uSerSerAspAspAspGluGlyGlnThrIleArgThrLeuGlyGluIleA 392

1347 ..... 1347
 392 rgSerGluLeuGluThrProLysProValGlnLeuGlnLeuProThrSer 408

1348 .....AGCACCAGCGG 1358
 409 SerThrThrThrProLeuLysThrThrSerGluAspAspSerThrSerVa 425

1359 GGAGCAGATGAAG...GTGTACAAGTGCGAACACTGCCGGGTGCTCTTCC 1405
 425 lGluProLeuGlnAsnLeuTyrGluCysLysPheCysAspIleSerPheL 442

1406 TGGATCACGTCATGTACACCATCCACATGGGCTGCCACGGCTTCCGTGAT 1455
 442 ysHisAlaValLeuTyrThrIleHisMetGlyTyrHisGlyTyrAsnAsp 458

1456 CCTTTTGAGTGCAACATGTGCGGCTACCACAGCCAGGACCGGTACGAGTT 1505
 459 ValPheLysCysAsnAlaCysGlyLysLysCysGluAspArgValAlaPh 475

1506 CTCGTCGCACATAACGCGAGGGGAGCAC 1533
 475 ePheLeuHisIleAlaArgAspAlaHis 484
```

```
seq_name: SwissProt_40:ZF37_HUMAN

seq_documentation_block:
ID   ZF37_HUMAN     STANDARD;      PRT;   597 AA.
AC   Q9Y6Q3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Zinc finger protein Zfp-37.
GN   ZFP37.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cartilage;
RX   MEDLINE=98252797; PubMed=9585434;
RA   Dreyer S.D., Zhou L., Machado M.A., Horton W.A., Zabel B.,
RA   Winterpacht A., Lee B.;
RT   "Cloning, characterization, and chromosomal assignment of the human
RT   ortholog of murine Zfp-37, a candidate gene for Nager syndrome.";
RL   Mamm. Genome 9:458-462(1998).
CC   -!- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVEL IN SEVERAL TISSUES
CC       INCLUDING FETAL HUMAN CARTILAGE.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF022158; AAC28425.1; ALT_INIT.
DR   HSSP; P08047; 1SP2.
DR   MIM; 602951; -.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR000822; Znf-C2H2.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 12.
DR   PRINTS; PR00048; ZINCFINGER.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 12.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 11.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 12.
KW   Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW   Nuclear protein; Repeat.
FT   DOMAIN        1     70       KRAB.
FT   DOMAIN      260    590       ZINC FINGERS.
FT   ZN_FING     260    282       C2H2-TYPE.
FT   ZN_FING     288    310       C2H2-TYPE.
```

```
FT   ZN_FING     316    334       C2H2-TYPE (ATYPICAL).
FT   ZN_FING     344    366       C2H2-TYPE.
FT   ZN_FING     372    394       C2H2-TYPE.
FT   ZN_FING     400    422       C2H2-TYPE.
FT   ZN_FING     428    450       C2H2-TYPE.
FT   ZN_FING     456    478       C2H2-TYPE.
FT   ZN_FING     484    506       C2H2-TYPE.
FT   ZN_FING     512    534       C2H2-TYPE.
FT   ZN_FING     540    562       C2H2-TYPE.
FT   ZN_FING     568    590       C2H2-TYPE.
SQ   SEQUENCE   597 AA;  67597 MW;  527B736119DFA446 CRC64;


alignment_scores:
            Quality:  331.50              Length:    537
              Ratio:  1.188                Gaps:     25
 Percent Similarity:  51.955   Percent Identity: 24.209

alignment_block:
 US-08-711-417C-165 x ZF37_HUMAN   ..

 Align seg 1/1  to: ZF37_HUMAN  from: 1  to: 597

      34 TCATCAGGGAAGGAAAGCCCCCCTGTAAGCGATACTCCAGATGAGGGCGA 83
         |||   ||||||                ||| ||||||::::::
     146 SerCysGlyLysIleLeuLysGlnAsnLeuAspLeuProAspHisSerAr 162

      84 TGAGCCCATGCCGATCCCCGAGGACCTCTCC..................... 114
          :::   :::          :::|||  :::
     162 gAsnCysValLysArgLysSerAspAlaAlaLysGluHisLysLysSerP 179

     115 ....................ACCACCTCGGGAGGACAGCAAAGCTCCAAG 144
                              |||  :::|||  :::|||::::::|||
     179 heAsnHisSerLeuSerAspThrArgLysGlyLysLysGlnThrGlyLys 195

     145 AGTGACAGAGTCGTGGCCAGTAATGTTAAAGTAGAGACTCAGAGTGATGA 194
                                  |||  |||     |||:::::
     196 ..........................LysHisGluLysLeuSerSerHi 203

     195 AGAGAATGGGCGTGCCTGTGAAATGAATGGGGAAGAA.............T 232
         ::::::::::      |||:::  :::|||::::::           |
     203 sSerSerSerAspLysCysAsnLysThrGlyLysLysHisAspLysLeuC 220

     233 GTGCGGAGGATTTACGAATGCTTGATGCCTCGGGAGAGAAAATGAATGGC 282
         ||:::::::::      :::     :::   :::|||::::::
     220 ysCysHisSerSerSerHisIle......LysGlnAspLysIleGlnThr 234

     283 TCCCACAGGGACCAAGGCAGCTCGGCTTTGTCGGGAGTTGGAGGCATTCG 332
         :::::::::  :::  |||  :::|||||||:::           ::
     235 GlyGluLysHisGluLysSerProSerLeuSerSerSerThrLysHisGl 251

     333 ACTTCCTAACGGAAAACTAAAG.........TGTGATATCTGTGGGATCA 373
         :  |||::::::  :::|||         |||:::  ||||||  :
     251 uLysProGlnAlaCysValLysProTyrGluCysAsnGlnCysGlyLysV 268

     374 TTTGCATCGGGCCCAATGTGCTCATGGTTCACAAAAGAAGCCACACTGGA 423
         ::           :::   |||:::  |||:::|||  |||||||||
     268 alLeuSerHisLysGlnGlyLeuIleAspHisGlnArgValHisThrGly 284

     424 GAACGGCCCTTCCAGTGCAATCAGTGCGGGGCCTCATTCACCCAGAAGGG 473
         |||:::|||::::::|||||:::|||||   :::|||:::|||||||::
     285 GluLysProTyrGluCysAsnGluCysGlyIleAlaPheSerGlnLysSe 301

     474 CAACCTGCTCCGGCACATCAAGCTGCATTCCGGGGAGAAGCCCTTCAAAT 523
         ::::|||:::   |||  :::  |||:::|||||||||||||::::::|
     301 rHisLeuValValHisGlnArgThrHisThrGlyGluLysProTyrGluC 318

     524 GCCACCTCTGCAACTACGCCTGCCGCCGGAGGGACGCCCTCACTGGCCAC 573
         ||     |||:::  |||      ::::::  ||||||||   |||
     318 ysIleGlnCysGlyLysAlaHisGlyHisLysHisAlaLeuThrAspHis 334

     574 CTGAGGACGCACTCCGTTGGTAAACCTCACAAATGTGGATATTGTGGCCG 623
         ||||||   |||:::      ||||||::::::|||:::   ||||||::

     335 LeuArgIleHisThrGlyGluLysProTyrGluCysAlaGluCysGlyLy 351

     624 AAGCTATAAACAGCGAACGTCTTTAGAGGAACATAAAGAGCGCTGCCACA 673
         ::::::::::::   ::::::|||  :::|||      |||  |||
     351 sThrPheArgHisSerSerAsnLeuIleGlnHis...ValArgSerHis. 366

     674 ACTACTTGGAAAGCATGGGCCTTCCGGGCACACTGTACCCAGTCATTAAA 723
         |||      |||  :::      :::  :::|||  :::|||::::::
     367 ThrGlyGluLysProTyrGluCysLys...GluCysGlyLysSerPheAr 382

     724 GAAGAAACTAAGCACAGTGAAATGGCAGAAGACCTGTGCAAGATAGGATC 773
         :  :::  |||  ::::::      :::|||  :::   :::|||
     382 gTyrAsnSerSerLeuThrGluHisValArgThrHisThr.GlyGluIle 398

     774 ...AGAGAGATCTCTCGTGCTGGACAGACTAGCAAGTAATGTCGCCAAAC 820
            |||  :::::::::|||::::::  |||
     399 ProTyrGluCysAsnGluCysGlyLysAlaPheLys.............. 410

     821 GTAAGAGCTCTATGCCTCAGAAATTTCTTGGGGACAAGGGCCTGTCCGAC 870
                 |||:::|||:::::::::
     411 ..........TyrSerSerSerLeuThr..................... 416

     871 ACGCCCTACGACAGTGCCACGTACGAGAAGGAGAACGAAATGATGAAGTC 920
                 ::::::  :::::::::  ||||||:::      |||
     417 ........LysHisMetArgIleHisThrGlyGluLysProPheGluCys 430

     921 CCACGTGATGGACCAAGC..........................CATCA 943
            ::   |||  :::                             |||||
     431 AsnGluCysGlyLysAlaPheSerLysLysSerHisLeuIleIleHisGl 447

     944 ACAACGCCATCAACTACCTGGGGGCCGAGTCCCTGCGCCCGCTGGTGCAG 993
         |:::  |||                          |||      ::::
     447 nArgThrHis..............ThrLysGluLysProTyrLysCysA 459

     994 ACGCCCCCGGGCGGTTCCGAGGTGGTCCCGGTCATCAGCCCGATGTACCA 1043
         ::      |||:::         :::  ||||||  |||
     459 snGluCysGlyLys.........AlaPheGlyHis.SerSerSerLeuTh 472

    1044 GCTGCACAGGCGCTCGGAGGGCACCCCGCGCTCCAACCACTCGGCCCAGG 1093
            |||  |||:::                    |||::::::::::
     472 rTyrHisMetArgThr.....................HisThrGlyGluS 482

    1094 ACAGCGCCGTGGAGTACCTGCTGCTGCTCTCCAAGGCCAAGTTGGTGCCC 1143
         ::                                          |||
     482 er.........................................Pro 483

    1144 TCGGAGCGCGAGGCGTCCCCGAGCAACAGCTGCCAAGACTCCACGGACAC 1193
            |||                  |||::|||
     484 PheGlu..............CysAsnGlnCys.................. 489

    1194 CGAGAGCAACAACGAGGAGCAGCGCAGCGGTCTTATCTACCTGACCAACC 1243
             :::::::::  :::|||  :::|||          ||||||:::|
     490 ....GlyLysGlyPheLysGlnIleGluGly.........LeuThrGlnH 502

    1244 ACATCGCCCGACGCGCGCAACGCGTGTCGCTCAAGGAGGAG...CACCGC 1290
         ||              |||||||||         |||:::  ::::::
     502 is..............GlnArgValHisThrGlyGluLysProTyrGlu 513

    1291 GCCTACGACCTGCTGCGCGCCGCCTCCGAGAACTCGCAGGACGCGCTCCG 1340
         :::  ::      :::|||  |||::::::|||:::      :::::
     514 CysAsnGluCysGlyLysAlaPheSerGlnLysSerHisLeuIleValHi 530

    1341 CGTGGTCAGCACCAGCGGGGAGCAGATGAAGGTGTACAAGTGCGAACACT 1390
         :      :::  :::|||||      |||  |||:::|||::::::|
     530 sGlnArgThrHisThrGlyGlu......LysProTyrGluCysAsnGluC 545

    1391 GCCGGGTGCTCTTCCTGGATCACGTCATGTACACCATCCACATGGGCTGC 1440
         ||:::  |||          :::  :::|||||:::
     545 ysGluLysAlaPheAsnAlaLysSerGlnLeuValIleHisGlnArgSer 561

    1441 CACGGCTTCCGTGATCCTTTTGAGTGCAACATGTGCGGCTACCACAGCCA 1490
         |||      :::  |||:::|||||||||  ||||||         ::
     562 HisThrGlyGluLysProTyrGluCysAsnGluCysGlyLysThrPheLy 578
```

```
     1491 GGACCGGTACGAGTTCTCGTCGCACATA...ACGCGAGGGGAGCACCGCT 1537
          :::::::   ::::::::::::|||:::   |||::::::|||   :::
      578 sGlnAsnAlaSerLeuThrLysHisValLysThrHisSerGluAspLysS 595

     1538 TCCAC 1542
            |||
      595 erHis 596

seq_name: SwissProt_40:ZF37_MOUSE

seq_documentation_block:
ID   ZF37_MOUSE     STANDARD;      PRT;   594 AA.
AC   P17141; Q62514;
DT   01-AUG-1990 (Rel. 15, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Zinc finger protein 37 (Zfp-37) (Male germ cell specific zinc finger
DE   protein).
GN   ZFP37 OR ZFP-37.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBA/CA, CBA/C57BL, AND C57BL/6; TISSUE=Brain;
RX   MEDLINE=96301401; PubMed=8660973;
RA   Mazarakis N., Michalovich D., Karis A., Grosveld F., Galjart N.;
RT   "Zfp-37 is a member of the KRAB zinc finger gene family and is
RT   expressed in neurons of the developing and adult CNS.";
RL   Genomics 33:247-257(1996).
RN   [2]
RP   SEQUENCE OF 42-594 FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=92310982; PubMed=1614869;
RA   Burke P.S., Wolgemuth D.J.;
RT   "Zfp-37, a new murine zinc finger encoding gene, is expressed in a
RT   developmentally regulated pattern in the male germ line.";
RL   Nucleic Acids Res. 20:2827-2834(1992).
RN   [3]
RP   SEQUENCE OF 219-594 FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=90301500; PubMed=2114017;
RA   Nelki D., Dudley K., Cunningham P., Akhavan M.;
RT   "Cloning and sequencing of a zinc finger cDNA expressed in mouse
RT   testis.";
RL   Nucleic Acids Res. 18:3655-3655(1990).
CC   -!- FUNCTION: MAY HAVE A ROLE IN REGULATING SPERMIOGENESIS.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN TESTIS AND BRAIN.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X89264; CAA61539.1; -.
DR   EMBL; X64413; CAA45758.1; -.
DR   EMBL; X52533; CAA36769.1; ALT_INIT.
DR   PIR; S10245; S10245.
DR   PIR; S22954; S22954.
DR   HSSP; P08047; 1SP2.
DR   MGD; MGI:99181; Zfp37.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR000822; Znf-C2H2.
DR   Pfam; PF00096; zf-C2H2; 12.
DR   PRINTS; PR00048; ZINCFINGER.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 12.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 11.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 12.
KW   Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW   Nuclear protein; Repeat; Developmental protein; Spermatogenesis.
FT   DOMAIN        3     74       KRAB.
FT   DOMAIN      255    585       ZINC FINGERS.
FT   ZN_FING     255    277       C2H2-TYPE.
FT   ZN_FING     283    305       C2H2-TYPE.
FT   ZN_FING     311    324       C2H2-TYPE.
FT   ZN_FING     339    361       C2H2-TYPE.
FT   ZN_FING     367    389       C2H2-TYPE.
FT   ZN_FING     395    417       C2H2-TYPE.
FT   ZN_FING     423    445       C2H2-TYPE.
FT   ZN_FING     451    473       C2H2-TYPE.
FT   ZN_FING     479    501       C2H2-TYPE.
FT   ZN_FING     507    529       C2H2-TYPE.
FT   ZN_FING     535    557       C2H2-TYPE.
FT   ZN_FING     563    585       C2H2-TYPE.
FT   CONFLICT    101    101       S -> T (IN REF. 1).
FT   CONFLICT    572    572       F -> V (IN REF. 3).
SQ   SEQUENCE   594 AA;  67254 MW;  B6CA2F9F53B70455 CRC64;

alignment_scores:
            Quality:  328.50              Length:    532
              Ratio:  1.226                 Gaps:     27
 Percent Similarity:  50.376   Percent Identity: 24.812

alignment_block:
 US-08-711-417C-165 x ZF37_MOUSE   ..

 Align seg 1/1  to: ZF37_MOUSE  from: 1  to: 594

      28 TCTTTCTCATCAGGGAAGGAAAGCCCCCCTGTAAGCGATACT..... 69
         |||   |||   ||||||:::|||         ||||||:::
     146 SerSerSerArgGlyLysAsnSerAsnGlnAsnSerAspSerLeuLysLy 162

      70 ....CCAGATGAGGGCGATGAG.......................... 87
             ||||||   ::::::|||
     162 sLysProAspThrAlaAsnGluHisArgLysSerLeuSerHisSerAlaS 179

      88 ................CCCATGCCGATCCCCGAGGACCTCTCCACCACC 120
                          :::|||      :::      :::
     179 erAspValAsnLysAspGluIleProThrArgLysLysCysAspLysLeu 195

     121 TCGGGAGGACAGCAAAGCTCCAAGAGTGACAGAGTCGTGGCCAGTAATGT 170
           :::::::::   |||:::|||:::|||:::
     196 ProAsnAsnLysLeuSerAspLysGlyAspLys.............. 206

     171 TAAAGTAGAGACTCAGAGTGATGAAGAGAATGGGCGTGCCTGTGAAATGA 220
                ::::|||:::::::::::      ::::::|||:::   :
     207 .........AsnGlnThrSerLysLysCysGluLysValCysArgHisS 220

     221 ATGGGGAAGAATGTGCGGAGGATTTACGAATGCTTGATGCCTCGGGAGAG 270
         ::::::::::      ||||||      :::          :::||||||
     220 erAlaSerHisThrLysGluAspLysIleGln.........ThrGlyGlu 233

     271 AAAATGAATGGCTCCCACAGGGACCAAGGCAGCTCGGCTTTGTCGGGAGT 320
         |||   :::   ||||||           |||:::      ::::::
     234 LysArgLys...SerHisCysArgThrProSerLysProGluLysAlaPr 249

     321 TGGAGGCATTCGACTTCCTAACGGAAAA...CTAAAGTGTGATATCTGTG 367
          |||              :::||||||      :::|||:::   ||||
     249 oGly..............SerGlyLysProTyrGluCysAsnHisCysG 261

     368 GGATCATTTGCATCGGGCCCAATGTGCTCATGGTTCACAAAAGAAGCCAC 417
         ||   :::           :::   |||:::   |||:::|||:::|||
     261 lyLysValLeuSerHisLysGlnGlyLeuLeuAspHisGlnArgThrHis 277
```

```
 418 ACTGGAGAACGGCCCTTCCAGTGCAATCAGTGCGGGGCCTCATTCACCCA 467
     |||||||||:::|||:::::|||||||:::||||||   :::|||:::||
 278 ThrGlyGluLysProTyrGluCysAsnGluCysGlyIleAlaPheSerGl 294

 468 GAAGGGCAACCTGCTCCGGCACATCAAGCTGCATTCCGGGGAGAAGCCCT 517
     ||||:::::::|||:::   |||   :::   |||:::||||||||||||:
 294 nLysSerHisLeuValLeuHisGlnArgThrHisThrGlyGluLysProT 311

 518 TCAAATGCCACCTCTGCAACTACGCCTGCCGCCGGAGGGACGCCCTCACT 567
     :::::|||:::   |||:::   |||      ::::::   |||||||||
 311 yrGluCysGluGlnCysGlyLysAlaHisGlyHisLysHisAlaLeuThr 327

 568 GGCCACCTGAGGACGCACTCCGTTGGTAAACCTCACAAATGTGGATATTG 617
       |||||||||   |||:::      ||||||:::||||||:::   ||
 328 AspHisLeuArgIleHisThrGlyGluLysProTyrLysCysAsnGluCy 344

 618 TGGCCGAAGCTATAAACAGCGAACGTCTTTAGAGGAACATAAAGAGCGCT 667
     ||||:::::::::::::::   ::::::|||   ::|||      |||
 344 sGlyLysThrPheArgHisSerSerAsnLeuMetGlnHisLeu...ArgS 360

 668 GCCACAACTACTTGGAAAGCATGGGCCTTCCGGGCACACTGTACCCAGTC 717
       ||| |||      |||   :::      :::   :::|||   :::||
 360 erHis.ThrGlyGluLysProTyrGluCysLys...GluCysGlyLysSe 375

 718 ATTAAAGAAGAAACTAAGCACAGTGAAATGGCAGAAGACCTGTGCAAGAT 767
     |::::::   :::   |||   ::::::      :::|||   :::
 375 rPheArgTyrAsnSerSerLeuThrGluHisValArgThrHisThr.Gly 391

 768 AGGATC...AGAGAGATCTCTCGTGCTGGACAGACTAGCAAGTAATGTCG 814
     :::|||      |||   ::::::::::|||::::::   |||
 392 GluIleProTyrGluCysAsnGluCysGlyLysAlaPheLys........ 405

 815 CCAAACGTAAGAGCTCTATGCCTCAGAAATTTCTTGGGGACAAGGGCCTG 864
                   |||:::|||:::::::::
 406 ................TyrGlySerSerLeuThr................ 411

 865 TCCGACACGCCCTACGACAGTGCCACGTACGAGAAGGAGAACGAAATGAT 914
                     ::::::   :::::::::   |||||||:::
 412 ..............LysHisMetArgIleHisThrGlyGluLysProPhe 423

 915 GAAGTCCCACGTGATGGACCAAGC........................... 938
     |||      :::   |||   :::
 424 GluCysAsnGluCysGlyLysThrPheSerLysLysSerHisLeuValIl 440

 939 .CATCAACAACGCCATCAACTACCTGGGGGCCGAGTCCCTGCGCCCGCTG 987
      |||||||:::   |||                              |||
 440 eHisGlnArgThrHis..............ThrLysGluLysProTyrL 452

 988 GTGCAGACGCCCCCGGGCGGTTCCGAGGTGGTCCCGGTCATCAGCCCGAT 1037
       :::|||      |||:::          :::   |||||| |||
 452 ysCysAspGluCysGlyLys.........AlaPheGlyHis.SerSerSe 465

1038 GTACCAGCTGCACAGGCGCTCGGAGGGCACCCCGCGCTCCAACCACTCGG 1087
               |||   |||:::                        |||::::
 465 rLeuThrTyrHisMetArgThr....................HisThrG 475

1088 CCCAGGACAGCGCCGTGGAGTACCTGCTGCTGCTCTCCAAGGCCAAGTTG 1137
     :::::
 475 lyAsp.......................................... 476

1138 GTGCCCTCGGAGCGCGAGGCGTCCCCGAGCAACAGCTGCCAAGACTCCAC 1187
        |||   |||                  |||:::|||      :::
 477 CysProPheGlu..............CysAsnGlnCysGlyLysAlaPh 488

1188 GGACACCGAGAGCAACAACGAGGAGCAGCGCAGCGGTCTTATCTACCTGA 1237
                            :::|||   :::|||         ||||
 488 e....................LysGlnIleGluGly.........LeuT 495

1238 CCAACCACATCGCCCGACGCGCGCAACGCGTGTCGCTCAAGGAGGAGCAC 1287
     ||:::|||                |||||||||         |||
 495 hrGlnHis..............GlnArgValHisThrGlyGlu...... 504

1288 CGCGCCTACGACCTGCTGCGCGCCGCCTCCGAGAACTCGCAGGACGCGCT 1337
```

```
          :::   |||:::   :::::::::::::::      ||||||   :::
 505 LysProTyrGluCysValGluCysGlyLysAlaPheSerGlnLysSerHi 521

1338 CCGCGTGGTC.........AGCACCAGCGGGGAGCAGATGAAGGTGTACA 1378
           :::|||         :::   :::|||||||      |||   ::::
 521 sLeuIleValHisGlnArgThrHisThrGlyGlu......LysProPheG 536

1379 AGTGCGAACACTGCCGGGTGCTCTTCCTGGATCACGTCATGTACACCATC 1428
     ::|||   :::|||         |||            :::   :::|||
 536 luCysTyrGluCysGlyLysAlaPheAsnAlaLysSerGlnLeuValIle 552

1429 CACATGGGCTGCCACGGCTTCCGTGATCCTTTTGAGTGCAACATGTGCGG 1478
     |||:::      |||      :::   |||:::||||||      |||||
 553 HisGlnArgSerHisThrGlyGluLysProTyrGluCysIleGluCysGl 569

1479 CTACCACAGCCAGGACCGGTACGAGTTCTCGTCGCACATA 1518
     |         :::::::::   :::::::::::::|||:::
 569 yLysAlaPheLysGlnAsnAlaSerLeuThrLysHisMet 582
```

```
seq_name: SwissProt_40:ZG20_XENLA

seq_documentation_block:
ID   ZG20_XENLA      STANDARD;      PRT;   675 AA.
AC   P18714;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Gastrula zinc finger protein XFG20-1 (XLCGF20.1).
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95094815; PubMed=8001572;
RA   Schaefer U., Rausch O., Bouwmeester T., Pieler T.;
RT   "Sequence-specific recognition of a repetitive DNA element by a C2H2
RT   zinc-finger protein in Xenopus.";
RL   Eur. J. Biochem. 226:567-576(1994).
RN   [2]
RP   SEQUENCE OF 85-613 FROM N.A.
RX   MEDLINE=90040698; PubMed=2509712;
RA   Nietfeld W., El-Baradi T., Mentzel H., Pieler T., Koester M.,
RA   Poeting A., Knoechel W.;
RT   "Second-order repeats in Xenopus laevis finger proteins.";
RL   J. Mol. Biol. 208:639-659(1989).
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X82643; CAA57965.1; -.
DR   PIR; S06565; S06565.
DR   HSSP; P25490; 1ZNM.
DR   TRANSFAC; T02366; -.
DR   InterPro; IPR000822; Znf-C2H2.
DR   Pfam; PF00096; zf-C2H2; 18.
DR   PRINTS; PR00048; ZINCFINGER.
DR   SMART; SM00355; ZnF_C2H2; 18.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 17.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 18.
KW   Zinc-finger; Metal-binding; DNA-binding; Repeat.
FT   ZN_FING      62     84       C2H2-TYPE.
FT   ZN_FING      90     112      C2H2-TYPE.
FT   ZN_FING      118    140      C2H2-TYPE.
FT   ZN_FING      146    168      C2H2-TYPE.
FT   ZN_FING      174    196      C2H2-TYPE.
FT   ZN_FING      202    224      C2H2-TYPE.
```

```
FT   ZN_FING     257    279       C2H2-TYPE.
FT   ZN_FING     286    308       C2H2-TYPE.
FT   ZN_FING     344    366       C2H2-TYPE.
FT   ZN_FING     373    395       C2H2-TYPE.
FT   ZN_FING     424    446       C2H2-TYPE.
FT   ZN_FING     452    474       C2H2-TYPE.
FT   ZN_FING     507    529       C2H2-TYPE.
FT   ZN_FING     535    557       C2H2-TYPE.
FT   ZN_FING     563    585       C2H2-TYPE.
FT   ZN_FING     591    613       C2H2-TYPE.
FT   ZN_FING     619    642       C2H2-TYPE.
SQ   SEQUENCE   675 AA;  77116 MW;  033094852C1FCF39 CRC64;
```

alignment_scores:

Quality:	323.00	Length:	619
Ratio:	1.099	Gaps:	21
Percent Similarity:	47.496	Percent Identity:	20.355

alignment_block:
US-08-711-417C-165 x ZG20_XENLA ..

Align seg 1/1 to: ZG20_XENLA from: 1 to: 675

```
      82 GATGAGCCCATGCCGATCCCCGAGGACCTCTCCACCACCTCG..... 123
         :::||||||     :::||||||:::         |||:::
       6 GluGluProTyrGluPheProGluAsnThrPheGlyThrAsnGluLeuLe 22

     124 .GGAGGACAGCAAAGCTCCAAGAGTGACAGAGTCGTGGCCAGTAATGTTA 172
           :::   |||:::::::   :::|||      :::   |||::::::|
      22 uProAsnTyrGlnGlnAsnCysThrAspGlyGluAlaIleSerAspThrL 39

     173 AAGTAGAG.....................ACTCAGAGTGATGAAGAG 198
         ||   :::                     |||::::::   ||||||
      39 ysSerAspLeuAlaTyrLeuGluValGluIleThrAspAlaHisGluGlu 55

     199 AATGGGCGTGCCTGTGAAATGAATGGGGAAGAATGTGCGGAGGATTTACG 248
         ::::::              ::::::      ||||||::::::   :::
      56 SerAsnThrAspLysProPheThrCysThrGluCysGlyLysThrPheTh 72

     249 AATGCTTGATGCCTCGGGAGAGAAAATGAAT...GGCTCCCACAGGGACC 295
                           :::|||   |||      ||||||
      72 r.................ArgLysProAsnTyrGluSerHis..... 80

     296 AAGGCAGCTCGGCTTTGTCGGGAGTTGGAGGCATTCGACTTCCTAACGGA 345
                                     ||||||     :::|||
      81 ............................IleArgAlaHisLysGly 86

     346 AAA......CTAAAGTGTGATATCTGTGGGATCATTTGCATCGGGCCCAA 389
         :::      ::::::|||   :::|||                    ::
      87 GluLysProPheSerCysMetValCysAspLysAlaPheAlaTrpLysSe 103

     390 TGTGCTCATGGTTCACAAAAGAAGCCACACTGGAGAACGGCCCTTCCAGT 439
         :   |||:::||||||         |||:::|||||||:::|||||||:::|
     103 rAsnLeuLeuValHisTyrSerValHisSerGlyGluLysProPheSerC 120

     440 GCAATCAGTGCGGGGCCTCATTCACCCAGAAGGGCAACCTGCTCCGGCAC 489
         ||::::::|||      :::|||:::::::|||:::::::|||   :::|||
     120 ysThrGluCysAspLysThrPheSerAsnLysAlaGlnLeuGluLysHis 136

     490 ATCAAGCTGCATTCCGGGGAGAAGCCCTTCAAATGCCACCTCTGCAACTA 539
         :::::::::|||:::|||||||||||:::::|||:::   |||:::
     137 LeuArgValHisThrGlyGluLysProTyrSerCysGluGlnCysGlyLy 153

     540 CGCCTGCCGCCGGAGGGACGCCCTCACTGGCCACCTGAGGACGCACTCCG 589
          :::      ::::::   :::|||   :::|||   ||||||||||:::
     153 sSerPheAlaHisLysCysValLeuAspSerHisGlnArgThrHisThrG 170

     590 TTGGTAAACCTCACAAATGTGGATATTGTGGCCGAAGCTATAAACAGCGA 639
             ||||||   :::|||      ||||||:::::::::::::||||||
     170 lyAspLysProPheSerCysThrGluCysGlyLysLysPheSerGlnArg 186

     640 ACGTCTTTAGAGGAACATAAAGAGCGC..................... 666
```

```
         :::|||:::::::|||   :::
     187 GlyAsnLeuHisLysHisLeuLysThrHisLysLeuAspGlnProHisLe 203

     667 .TGCCACAACTACTTGGAAAGCATGGGCCTTCCGGGCACACTGTACCCAG 715
          |||   :::      :::::::::::::::   :::||||||
     203 uCysAlaGluCysGlyLysThrPheSerPheLysSerThrLeu..... 217

     716 TCATTAAAGAAGAAACTAAGCACAGTGAAATG................. 747
           ::::::::::      ||||||||||:::
     218 ..LeuGluHisGlnLysIleHisSerGluIleArgProLeuSerGluPhe 233

     748 .................GCAGAAGACCTGTGCAAG.............. 765
                         |||::::::|||   |||
     234 GlyLysThrPheSerAspAlaHisAsnLeuLeuLysHisGlnSerThrPh 250

     766 ............................................ATAG 769
                                                     ::::
     250 eThrGluGluGlnLysProPheProCysThrGluCysGlyGluIlePheS 267

     770 GATCAGAGAGATCTCTCGTGCTGGACAGA.................... 798
         :::::|||::::::|||:::     :::
     267 erAsnGluHisGluLeuLeuThrHisGlnSerThrHisThrGluGluGln 283

     799 ..........................CTAGCAAGTAATGTCGCCAA 818
                                     :::   ||||||      ::
     284 LysProPheProCysThrLysCysTrpGlyIlePheSerAsnGluHisGl 300

     819 ACGTAAGAGCTCTATGCCTCAGAAATTTCTTGGGGACAAGGGCCTGTCCG 868
         :   :::::   :::               |||:::|||:::|||   :
     300 uLeuArgThrHisGlnSerThrHisThrGluGlyGlnLysSerLeuProS 317

     869 ACACGCCCTACGACAGTGCCACGTACGAGAAGGAGAACGAAATGATG... 915
         ::|||      :::::::::|||:::::::::|||:::|||::::::
     317 erThrGlu...SerGlyGlyThrPheSerAsnGluHisGluLeuLeuThr 332

     916 ......AAGTCCCACGTGATGGACCAA..................... 936
                ::::::|||:::   :::|||
     333 HisGlnSerThrHisThrGluGluGlnLysHisLeuProCysThrGluCy 349

     937 .......GCCATCAACAACGCCATCAACTACCTGGGGGCCGAGTCCCTGC 979
                 ::::::::::|||      :::   |||:::   :::|||   :
     349 sGlyGlyThrPheThrAsnGluGlnGluLeuLeuAlaHisGlnSerThrH 366

     980 GCCCGCTGGTGCAGACGCCCCCGGGCGGTTCCGAGGTGGTCCCGGTCATC 1029
         ::         |||   |||         :::|||         ::::::
     366 isThrGluGluGlnLysProLeuProCysThrGluCysGlyGluIlePhe 382

    1030 AGCCCGATGTACCAGCTG......CACAGGCGCTCGGAGGGCACCCCGCG 1073
         |||      ::::::|||      |||:::   ::::::   :::|||
     383 SerAspGluHisGluLeuLeuThrHisGlnSerThrHisThrSerProSe 399

    1074 CTCCAACCACTCGGCCCAGGACAGCGCCGTGGAGTACCTGCTGCTGCTCT 1123
          ::::::   :::::::|||  :::
     399 rThrGluPheGlyValGlnThrThr..................... 407

    1124 CCAAGGCCAAGTTGGTGCCCTCGGAGCGCGAGGCGTCCCCGAGCAAC... 1170
                           ::::::::::::   ||||||||||:::
     408 ...................GluAspAsnHisGlnSerProSerLysAsp 417

    1171 ....................AGCTGCCAAGACTCCACGGACACCGAGAG 1199
                             ||||||::::::         :::
     418 HisThrGlyGluLysProPheSerCysSerGluCysGlyLysSerPhePh 434

    1200 CAACAACGAGGAGCAGCGCAGCGGTCTTATCTACCTGACCAACCACATCG 1249
          ::::::      ::::::   |||:::          |||   :
     434 eTyrLysSerValLeuLysAspHisLeuValVal.....HisThrG 448

    1250 CCCGACGC..............................GCGCAACGC 1266
         ::::::::                                ::::::::
     448 lyLysLysProTyrHisCysIleGluCysGlyArgSerTyrThrHisGln 464

    1267 GTGTCGCTCAAGGAGGAGCACCGCGCCTACGACCTGCTGCGCGCCGCCTC 1316
         |||||||||:::::::::|||::::::      ::::::|||   ||
```

```
     465 SerSerLeuLysSerHisGlnArgThrHisThrGlyValLysAlaPheSe 481

    1317 CGAGAACTCGCAGGACGCGCTCCGCGTGGTCAGC..................  1350
         |   |||       |||   |||   ::::::|||
     481 rCysAsnLeuCysAspLysLeuSerIleIleSerLysLeuArgLeuHisT 498

    1350 ..................................................  1350

     498 yrArgValHisSerGlyGluLysProTyrProCysThrGluCysAspLys 514

    1351 ......ACCAGCGGGGAGCAGATG.........................  1368
               |||:::   ||||||:::
     515 ThrPheThrLysLysGluGlnLeuGluSerHisTyrLysValHisThrGl 531

    1369 ....AAGGTGTACAAGTGCGAACACTGCCGGGTGCTCTTCCTGGATCACG 1414
             |||   |||   |||:::::::|||          |||
     531 yGluLysProTyrProCysGlnGlnCysGlyLysSerPheSerHisLysS 548

    1415 TCATGTACACCATCCACATGGGCTGCCACGGCTTCCGTGATCCTTTTGAG 1464
           :::      :::|||:::      |||             ||||||:::
     548 erValLeuLysLeuHisLeuArgThrHisThrGlyAspLysProPheSer 564

    1465 TGCAACATGTGCGGCTACCACAGCCAGGACCGGTACGAGTTCTCGTCGCA 1514
         |||:::   ||||||              :::   :::::::::||||||
     565 CysThrGluCysGlyLysThrPheThrArgLysProAsnTyrGluSerHi 581

    1515 CATAACG 1521
         |:::|||
     581 sLeuThr 583
```

```
seq_name: SwissProt_40:Z184_HUMAN

seq_documentation_block:
ID   Z184_HUMAN     STANDARD;      PRT;   751 AA.
AC   Q99676; O60792;
DT   15-JUL-1998 (Rel. 36, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Zinc finger protein 184.
GN   ZNF184.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Phillips S.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 26-751 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=97230463; PubMed=9073517;
RA   Goldwurm S., Menzies M.L., Banyer J.L., Powell B.L.W.,
RA   Jazwinska E.C.;
RT   "Identification of a novel Krueppel-related zinc finger gene (ZNF184)
RT   mapping to 6p21.3.";
RL   Genomics 40:486-489(1997).
CC   -!- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: PREDOMINANT EXPRESSION IN TESTIS.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL021918; CAA17278.1; -.
DR   EMBL; U66561; AAC51180.1; -.
DR   HSSP; P08046; 1A1H.
DR   MIM; 602277; -.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR000822; Znf-C2H2.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 19.
DR   PRINTS; PR00048; ZINCFINGER.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 19.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 19.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 19.
KW   Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW   Nuclear protein; Repeat.
FT   DOMAIN       28     99       KRAB.
FT   DOMAIN      222    748       ZINC FINGERS.
FT   ZN_FING     222    244       C2H2-TYPE.
FT   ZN_FING     250    272       C2H2-TYPE.
FT   ZN_FING     278    300       C2H2-TYPE.
FT   ZN_FING     306    328       C2H2-TYPE.
FT   ZN_FING     334    356       C2H2-TYPE.
FT   ZN_FING     362    384       C2H2-TYPE.
FT   ZN_FING     390    412       C2H2-TYPE.
FT   ZN_FING     418    440       C2H2-TYPE.
FT   ZN_FING     446    468       C2H2-TYPE.
FT   ZN_FING     474    496       C2H2-TYPE.
FT   ZN_FING     502    524       C2H2-TYPE.
FT   ZN_FING     530    552       C2H2-TYPE.
FT   ZN_FING     558    580       C2H2-TYPE.
FT   ZN_FING     586    608       C2H2-TYPE.
FT   ZN_FING     614    636       C2H2-TYPE.
FT   ZN_FING     642    664       C2H2-TYPE.
FT   ZN_FING     670    692       C2H2-TYPE.
FT   ZN_FING     698    720       C2H2-TYPE.
FT   ZN_FING     726    748       C2H2-TYPE.
FT   CONFLICT     27     27       A -> S (IN REF. 2).
FT   CONFLICT    295    295       Q -> R (IN REF. 2).
FT   CONFLICT    359    359       E -> G (IN REF. 2).
FT   CONFLICT    379    381       QHQ -> PHP (IN REF. 2).
FT   CONFLICT    498    498       R -> G (IN REF. 2).
FT   CONFLICT    526    526       Q -> G (IN REF. 2).
FT   CONFLICT    533    533       K -> E (IN REF. 2).
FT   CONFLICT    547    549       KHE -> QHQ (IN REF. 2).
FT   CONFLICT    746    746       R -> K (IN REF. 2).
SQ   SEQUENCE   751 AA;  86174 MW;  249DFB302C6BC97D CRC64;


alignment_scores:
            Quality:  323.00              Length:    418
              Ratio:  1.455                 Gaps:     18
 Percent Similarity:  53.110   Percent Identity: 25.359

alignment_block:
 US-08-711-417C-165 x Z184_HUMAN   ..

 Align seg 1/1  to: Z184_HUMAN  from: 1  to: 751

     352 AAGTGTGATATCTGTGGGATCATTTGCATCGGGCCCAATGTGCTCATGGT 401
         |||||||:::   ||||||            ||||||:::::::::::::
     391 LysCysAsnGluCysGlyLysAlaPheAsnGlyProSerThrPheIleAr 407

     402 TCACAAAAGAAGCCACACTGGAGAACGGCCCTTCCAGTGCAATCAGTGCG 451
          |||          ||||||||||||:::|||::::::||||||:::||||
     407 gHisHisMetIleHisThrGlyGluLysProTyrGluCysAsnGluCysG 424

     452 GGGCCTCATTCACCCAGAAGGGCAACCTGCTCCGGCACATCAAGCTGCAT 501
         ||   :::|||:::|||   :::||||||   :::|||   |||   |||
     424 lyLysAlaPheSerGlnHisSerAsnLeuThrGlnHisGlnLysThrHis 440

     502 TCCGGGGAGAAGCCCTTCAAATGCCACCTCTGCAACTACGCCTGCCGCCG 551
         :::||||||||||||:::   |||      |||:::   :::
     441 ThrGlyGluLysProTyrAspCysAlaGluCysGlyLysSerPheSerTy 457
```

```
 552 GAGGGACGCCCTCACTGGCCACCTGAGGACGCACTCCGTTGGTAAACCTC 601
     ::::::|||:::   ||||||:::   |||:::      ||||||:
 457 rTrpSerSerLeuAlaGlnHisLeuLysIleHisThrGlyGluLysProT 474

 602 ACAAATGTGGATATTGTGGCCGAAGCTATAAACAGCGAACGTCTTTAGAG 651
     ::|||||||:::   ||||||:::::::::::::      :::||||||
 474 yrLysCysAsnGluCysGlyLysAlaPheSerTyrCysSerSerLeuThr 490

 652 GAACATAAAGAG.....................CGCTGCCACAACTA 677
     :::|||::::::                     :::||| :::
 491 GlnHisArgArgIleHisThrArgGluLysProPheGluCys.SerGluC 507

 678 CTTGGAAAGCATGGGCCTTCCGGGCACACTGTACCCAGTCATTAAAGAAG 727
         ||||||:::::::::   ::::::   :::|||      |||
 507 ysGlyLysAlaPheSerTyrLeuSerAsnLeuAsnGlnHisGlnLysThr 523

 728 AAACTAAGCACAGTGAAATGGCAG...AAGACCTGTGCAAGATAGG...A 771
         :::        ::::::   |||   |||::::::       |
 524 HisThrGlnGluLysAlaTyrGluCysLysGluCysGlyLys.AlaPheI 540

 772 TCAGAGAGATCTCTCGTGCT.....................GGACAGACT 800
     |||||:::   |||   |||                     |||:::
 540 leArgSerSerSerLeuAlaLysHisGluArgIleHisThrGlyGluLys 556

 801 AGCAAGTAATGTCGCCAAACGTAAGAGCTCTATGCCTCAGAAATTTCTTG 850
           |||::::::                 ::::::   :::|||::
 557 ProTyrGlnCysHisGluCys............GlyLysThrPheSerTy 569

 851 GGGACAAGGGCCTGTCCGACACGCCCTACGACAGTGCCACGTACGAGAAG 900
     :|||::::::   ::::::|||     |||:::   |||:::      |
 569 rGlySerSerLeuIleGlnHis......ArgLysIleHisThr......G 582

 901 GAGAACGAAATGATGAAGTCCCACGTGATGGACCAAGCCATCAACAACGC 950
     ||||||||      |||   :::          :::|||:::|||:::
 582 lyGluArg.ProTyrLysCysAsnGluCysGlyArgAlaPheAsnGlnAs 598

 951 CATCAACTACCTGGGGGCCGAGTCCCTGCGCCCGCTGGTGCAGACGCCCC 1000
      |||:::                          |||:::|||
 598 nIleHis.........................LeuThrGln....... 603

1001 CGGGCGGTTCCGAGGTGGTCCCGGTCATCAGCCCGATGTACCAGCTGCAC 1050
                                                    |||
 604 ..............................................His 604

1051 AGGCGCTCGGAGGGCACCCCGCGCTCCAACCACTCGGCCCAG...GACAG 1097
     :::|||   :::          :::      :::   |||:::      ::
 605 LysArgIleHisThrGlyAlaLysProTyrGluCysAlaGluCysGlyLy 621

1098 CGCCGTGGAGTACCTGCTGCTGCTCTCCAAGGCCAAGTTGGTGCCCTCGG 1147
     :|||   ::::::         |||::::::   :::   :::   :::|
 621 sAlaPheArgHisCysSerSerLeuAlaGlnHisGlnLysThrHisThrG 638

1148 AGCGCGAGGCGTCCCCGAGCAACAGCTGCCAAGACTCCACGGACACCGAG 1197
     ||::::::            |||:::|||:::   :::   :::   :::
 638 luGluLysProTyrGlnCysAsnLysCysGluLysThrPheSerGlnSer 654

1198 AGCAACAACGAGGAGCAGCGCAGCGGTCTTATCTACCTGACCAACCACAT 1247
     |||                          :::||||||:::|||
 655 Ser............................HisLeuThrGlnHis.. 660

1248 CGCCCGACGCGCGCAACGCGTGTCGCTCAAGGAGGAG...CACCGCGCCT 1294
                 ||||||:::         |||:::   :::::::::
 661 .............GlnArgIleHisThrGlyGluLysProTyrLysCysA 673

1295 ACGACCTGCTGCGCGCCGCCTCCGAGAACTCGCAGGACGCGCTCCGCGTG 1344
       :::      :::|||   |||:::::::::::::   :::   :::
 673 snGluCysAspLysAlaPheSerArgSerThrHisLeuThrGluHisGln 689

1345 GTCAGCACCAGCGGGGAGCAGATGAAGGTGTACAAGTGCGAACACTGCCG 1394
       :::   :::||||||      |||   |||:::|||::::::|||||
 690 AsnThrHisThrGlyGlu......LysProTyrAsnCysAsnGluCysAr 704

1395 GGTGCTCTTCCTGGATCACGTCATGTACACCATCCACATGGGCTGCCACG 1444
     |      |||   :::   :::            |||:::      |||:
 704 gLysThrPheSerGlnSerThrTyrLeuIleGlnHisGlnArgIleHisS 721

1445 GCTTCCGTGATCCTTTTGAGTGCAACATGTGCGGCTACCACAGCCAGGAC 1494
     ::   :::   ||||||   ||||||   ||||||         :::
 721 erGlyGluLysProPheGlyCysAsnAspCysGlyLysSerPheArgTyr 737

1495 CGGTACGAGTTCTCGTCGCACATAACGCGAGGGGAGCACCGCTTCCAC 1542
     |||      :::::::::|||                ::::|||:::|||
 738 ArgSerAlaLeuAsnLysHis..............GlnArgLeuHis 748
```

```
seq_name: SwissProt_40:HUNB_TRICA

seq_documentation_block:
ID   HUNB_TRICA     STANDARD;      PRT;   524 AA.
AC   Q01791;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hunchback protein.
GN   HB.
OS   Tribolium castaneum (Red flour beetle).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC   Cucujiformia; Tenebrionidae; Tribolium.
OX   NCBI_TaxID=7070;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96125163; PubMed=8575322;
RA   Wolff C., Sommer R., Schroeder R., Glaser G., Tautz D.;
RT   "Conserved and divergent expression aspects of the Drosophila
RT   segmentation gene hunchback in the short germ band embryo of the
RT   flour beetle Tribolium.";
RL   Development 121:4227-4236(1995).
RN   [2]
RP   SEQUENCE OF 243-311 FROM N.A.
RX   MEDLINE=93066327; PubMed=1438276;
RA   Sommer R.J., Retzlaff M., Goerlich K., Sander K., Tautz D.;
RT   "Evolutionary conservation pattern of zinc-finger domains of
RT   Drosophila segmentation genes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:10782-10786(1992).
CC   -!- FUNCTION: GAP CLASS SEGMENTATION PROTEIN THAT CONTROLS DEVELOPMENT
CC       OF HEAD STRUCTURES.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: BELONGS TO THE HUNCHBACK FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X91618; CAA62821.1; -.
DR   EMBL; L01615; AAA30095.1; -.
DR   HSSP; P08151; 2GLI.
DR   InterPro; IPR000822; Znf-C2H2.
DR   Pfam; PF00096; zf-C2H2; 6.
DR   SMART; SM00355; ZnF_C2H2; 6.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 2.
KW   Developmental protein; Gap protein; Zinc-finger;
KW   Metal-binding; DNA-binding; Repeat; Nuclear protein.
FT   DOMAIN      202    311       ZINC FINGERS I.
FT   DOMAIN      471    523       ZINC FINGERS II.
FT   ZN_FING     202    224       C2H2-TYPE.
FT   ZN_FING     231    253       C2H2-TYPE.
FT   ZN_FING     259    281       C2H2-TYPE.
FT   ZN_FING     298    311       C2H2-TYPE.
FT   ZN_FING     471    493       C2H2-TYPE.
```

```
FT   ZN_FING     499    523       C2H2-TYPE.
SQ   SEQUENCE   524 AA;  59514 MW;  022EE44111BE4B9F CRC64;


alignment_scores:
            Quality:  317.00              Length:    561
              Ratio:  1.157                Gaps:     23
 Percent Similarity:  48.841   Percent Identity: 24.242

alignment_block:
 US-08-711-417C-165 x HUNB_TRICA   ..

 Align seg 1/1  to: HUNB_TRICA  from: 1  to: 524

      1 ATGGATGCTGACGAGGGTCAAGACATGTCTTTCTCATCAGGGAAGGAA.. 48
        ||||||  |||:::  :::|||  :::  :::||||||::::::
     48 MetAspGluAspLys...AsnAspSerGlyValThrSerGlySerAspPh 63

     49 .......AGCCCCCCTGTAAGCGATACTCCAGATGAG..... 78
                |||  |||  |||||||||  ::::::
     63 eHisSerSerSerProSerSerAspThrSerGlnAspLeuGlnHisSerT 80

     79 ..........GGCGATGAGCCC.............ATGCCGATCCCC 102
                      :::|||                  ||||||
     80 yrGlnSerProGlnThrGlnProAlaArgPheTyrSerThrProIleVal 96

    103 GAGGACCTCTCCACCACCTCGGGAGGACAGCAAAGCTCCAAGAGTGACAG 152
          ::::::  :::  :::                 :::|||:::
     97 ProHisPheAlaTyrAsnHisAsnProLeuThrProProAsnSerGluPr 113

    153 AGTCGTGGCCAGTAATGTTAAAGTAGAGACTCAGAGTGATGAAGAGAATG 202
         :::|||:::  :::  :::  |||  ::::::  |||  :::
    113 oLeuValSerProLysSerGluLysGluGluLysAspMetGluThrThrL 130

    203 GGCGTGCCTGT.........GAAATGAATGGGGAAGAATGTGCGGAGGAT 243
              |||         :::  :::  ::::::      :::
    130 euThrProCysAlaSerProAsnArgLysProAspAspAsnGlnAspHis 146

    244 TTACGAATGCTTGATGCCTCGGGAGAGAAA.........ATGAATGGCTC 284
        ||||||  |||:::  |||  ||||||          ::::::::::
    147 LeuArgArgLeuGluMetSerLeuGluLysSerGlyLeuPheSerSerLy 163

    285 CCACAGGGACCAAGGCAGCTCGGCTTTGTCGGGAGTTGGA......... 324
        :      :::::::::  :::  |||||||||  :::
    163 sThrSerGluHisSerValAspGluLeuSerGlyLysSerAspAsnAspA 180

    325 ...................GGCATTCGACTTCCT............AAC 342
                           ::::::|||:::|||           :::
    180 laGluGluTyrAspGluGlnSerLeuArgValProLysValAsnSerHis 196

    343 GGAAAACTA.........AAGTGTGATATCTGTGGGATCATTTGCATCGG 383
        |||||||:::         ||||||      |||  :::::::::|||
    197 GlyLysIleLysThrPheLysCysLysGlnCysAspPheValAlaIleTh 213

    384 GCCCAATGTGCTCATGGTTCACAAAAGAAGCCACACTGGAGAA...CGGC 430
                        |||::::::  |||      |||  :::
    213 rLysLeuGluGlnTrpAsnHisSerLysValHisIleArgGluAspLysA 230

    431 CCTTCCAGTGCAATCAGTGCGGGGCCTCATTCACCCAGAAGGGCAACCTG 480
          :::  |||  :::|||                  |||  :::|||
    230 rgLeuThrCysProLysCysProPheIleThrGluTyrLysHisHisLeu 246

    481 CTCCGGCACATCAAGCTGCATTCCGGGGAGAAGCCCTTCAAATGCCACCT 530
              |||::::::  |||:::|||:::||||||||||:::|||:::
    247 GluTyrHisLeuArgAsnHisAlaGlySerLysProPheGlnCysAsnLy 263

    531 CTGCAACTACGCCTGCCGCCGGAGGGACGCCCTCACTGGCCACCTGAGGA 580
         |||:::|||:::|||  :::::::::  |||::::::|||:::::::
    263 sCysAspTyrThrCysValAsnLysSerMetLeuAsnSerHisMetLysS 280

    581 CGCACTCCGTTGGTAAACCTCACAAATGTGGATATTGTGGCCGAAGCTAT 630
        ::||||||                ::::::|||      |||:::  :::
    280 erHisSerAsnValTyrArgTyrSerCysArgAspCysSerTyrAlaThr 296

    631 AAACAGCGAACGTCTTTAGAGGAACATAAAGAGCGCTGCCACAACTACTT 680
        |||         ||||||:::  |||
    297 LysTyrCysHisSerLeuLysIleHis..................... 305

    681 GGAAAGCATGGGCCTTCCGGGCACACTGTACCCAGTCATTAAAGAAGAAA 730

    305 ..................................................... 305

    731 CTAAGCACAGTGAAATGGCAGAAGACCTGTGCAAGATAGGATCAGAGAGA 780
                                    |||  :::  |||
    306 ......................LeuArgArgTyrGlyHisThrPro 313

    781 TCTCTCGTGCTGGACAGACTAGCAAGTAATGTCGCCAAACGTAAGAGCTC 830
        ::::::|||||||||:::  ::::::       |||::::::
    314 AsnValValLeuAspGluGluGlyAsn.ProCysProAspIle..... 327

    831 TATGCCTCAGAAATTTCTTGGGGACAAGGGCCTGTCCGACACGCCCTACG 880
          :::  :::      |||||||||      |||         |||
    328 .....IleIleAspValHisGlyThrArgArgGlyProLysIleLysThr 342

    881 ACAGTGCCACGTACGAGAAGGAGAACGAAATGATGAAGTCCCACGTGATG 930
        |||      ::::::|||::::::              |||:::
    343 .GlnProLysAlaGluGluAlaLys.............ProGluThrL 354

    931 GACCAAGCCATCAACAACGCCATCAACTACCTGGGGGCCGAGTCCCTGCG 980
        |||      :::  ::::::|||||||||            |||
    354 euProPheLeuAsnLeuGlnGlnGlnLeuProPheProGlyTyrPro... 369

    981 CCCGCTGGTGCAGACGCCCCCGGGCGGTTCCGAGGTGGTCCCGGTCATCA 1030
                                  |||  ||||||
    370 ........................PhePheGlyGly........... 373

   1031 GCCCGATGTACCAGCTGCACAGGCGCTCGGAGGGCACCCCGCGCTCCAAC 1080
              |||  ||||||  ||| :::         |||::::::
    374 .......PheProAsnAlaGlnLeuLeu.GlnGlnLeuIleArgGluArg 387

   1081 CACTCGGCCCAGGACAGCGCCGTGGAGTACCTGCTGCTGCTCTCCAAGGC 1130
        :::  |||      ::::::  |||         :::|||:::  ::
    388 GlnLeuAlaValGlyGlySerGlnGluGluSerArgValLeuAspLeuSe 404

   1131 CAAGTTGGTGCCC.....................TCGGAGCGCGAGG 1156
        :|||      |||                     ::::::|||:::
    404 rLys......ProGlyCysSerTyrThrGlyGluGlnLysSerArgArgL 419

   1157 CGTCCCCGAGCAACAGCTGCCAAGACTCCACGGACACCGAGAGCAACAAC 1206
          :::|||:::  :::      |||  |||::::::||||||::::::
    419 ysGlyProAlaPheLysVal...AspProThrGlnValGluSerGluGlu 434

   1207 GAGGAGCAGCGCAGCGGTCTTATCTACCTGACCAACCACATCGCCCGACG 1256
        |||:::::::::::::::         :::::::|||      ::::::
    435 GluAspGluGluThrSerThrThrValPheSerAsn......ValGluVa 449

   1257 CGCGCAACGCGTGTCGCTCAAGGAGGAGCACCGCGCCTACGACCTGCTGC 1306
          :::|||:::  :::  ||||||||      :::  :::      :
    449 lValGlnGluGluAlaLysLysGluGluSerAspSerAsnAsnAsnAsnA 466

   1307 GCGCCGCCTCCGAGAACTCGCAGGACGCGCTCCGCGTGGTCAGCACCAGC 1356
        ::      :::  ||||||
    466 snLysGluGluGlyAsnSer............................ 472

   1357 GGGGAGCAGATGAAGGTGTACAAGTGCGAACACTGCCGGGTGCTCTTCCT 1406
                               |||::::::|||::::::   |||
    473 ......................CysGlnTyrCysAsnIleAlaPheGl 481

   1407 GGATCACGTCATGTACACCATCCACATGGGCTGCCACGGCTTCCGTGATC 1456
        |||  |||:::||||||||||||||||  ||||||||:::::::|
    481 yAspAlaValLeuTyrThrIleHisMetGlyTyrHisGlyPheHisAsnP 498

   1457 CTTTTGAGTGCAACATGTGCGGCTACCACAGCCAGGACCGGTACGAGTTC 1506
        |||||  ||||||||||||||  :::  :::|||:::  :::|||
    498 roPheThrCysAsnMetCysGlyValGluCysSerAspLysValSerPhe 514
```

```
    1507 TCGTCGCACATAACGCGAGGGGAGCAC 1533
                 ||||||:::|||   :::|||
     515 PheLeuHisIleAlaArgValSerHis 523

seq_name: SwissProt_40:YD49_HUMAN

seq_documentation_block:
ID   YD49_HUMAN     STANDARD;      PRT;   739 AA.
AC   Q9P2J8;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Hypothetical zinc finger protein KIAA1349.
GN   KIAA1349.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=20181126; PubMed=10718198;
RA   Nagase T., Kikuno R., Ishikawa K., Hirosawa M., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XVI.
RT   The complete sequences of 150 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 7:65-73(2000).
CC   -!- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB037770; BAA92587.1; ALT_INIT.
DR   HSSP; P08151; 2GLI.
DR   InterPro; IPR000822; Znf-C2H2.
DR   Pfam; PF00096; zf-C2H2; 20.
DR   PRINTS; PR00048; ZINCFINGER.
DR   SMART; SM00355; ZnF_C2H2; 21.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 20.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 21.
KW   Hypothetical protein; Transcription regulation; DNA-binding;
KW   Zinc-finger; Metal-binding; Nuclear protein; Repeat.
FT   DOMAIN      587    732       ZINC FINGERS.
FT   ZN_FING     150    172       C2H2-TYPE.
FT   ZN_FING     178    200       C2H2-TYPE.
FT   ZN_FING     206    228       C2H2-TYPE.
FT   ZN_FING     234    256       C2H2-TYPE (DEGENERATE).
FT   ZN_FING     262    284       C2H2-TYPE.
FT   ZN_FING     290    312       C2H2-TYPE.
FT   ZN_FING     318    340       C2H2-TYPE.
FT   ZN_FING     346    368       C2H2-TYPE.
FT   ZN_FING     374    396       C2H2-TYPE.
FT   ZN_FING     402    424       C2H2-TYPE.
FT   ZN_FING     430    452       C2H2-TYPE.
FT   ZN_FING     458    480       C2H2-TYPE.
FT   ZN_FING     486    508       C2H2-TYPE.
FT   ZN_FING     514    536       C2H2-TYPE.
FT   ZN_FING     542    564       C2H2-TYPE.
FT   ZN_FING     570    592       C2H2-TYPE.
FT   ZN_FING     598    620       C2H2-TYPE.
FT   ZN_FING     626    648       C2H2-TYPE.
FT   ZN_FING     654    676       C2H2-TYPE.
FT   ZN_FING     682    704       C2H2-TYPE.
FT   ZN_FING     710    732       C2H2-TYPE.
SQ   SEQUENCE   739 AA;  85592 MW;  22617253E4C2B6CF CRC64;
```

```
alignment_scores:
            Quality:  315.00              Length:    453
              Ratio:  1.529                 Gaps:     15
 Percent Similarity:  45.475   Percent Identity: 24.283

alignment_block:
 US-08-711-417C-165 x YD49_HUMAN   ..

 Align seg 1/1  to: YD49_HUMAN  from: 1  to: 739

     352 AAGTGTGATATCTGTGGGATCATTTGCATCGGGCCCAATGTGCTCATGGT 401
         ||||||:::   ||||||         |||:::   :::   ||||||||
     207 LysCysAsnGluCysGlyLysAlaPheIleAlaSerSerSerLeuMetVa 223

     402 TCACAAAAGAAGCCACACTGGAGAACGGCCCTTCCAGTGCAATCAGTGCG 451
         ||||:::|||   ||||||   |||:::|||:::|||||||||   ||||
     223 lHisGlnArgIleHisThrLysGluLysProTyrGlnCysAsnValCysG 240

     452 GGGCCTCATTCACCCAG..................................... 468
         ||   ||||||:::|||
     240 lyLysSerPheSerGlnCysAlaArgLeuAsnGlnHisGlnArgIleGln 256

     468 ..................................................... 468

     257 ThrGlyGluLysProTyrLysCysSerGluCysGlyLysAlaPheSerAs 273

     469 .AAGGGCAACCTGCTCCGGCACATCAAGCTGCATTCCGGGGAGAAGCCCT 517
          |||::::::|||  ||||||   :::   |||:::|||||||||||:
     273 pLysSerLysLeuAlaArgHisGlnGluThrHisAsnGlyGluLysProT 290

     518 TCAAATGCCACCTCTGCAACTACGCCTGCCGCCGGAGGGACGCCCTCACT 567
         ::||||||      |||:::   |||   |||:::::::::   |||:::
     290 yrLysCysAspAspCysGlyLysAlaPheArgAsnLysSerTyrLeuSer 306

     568 GGCCACCTGAGGACGCACTCCGTTGGTAAACCTCACAAATGTGGATATTG 617
            |||   :::||||||:::      ||||||:::::|||:::   ||
     307 ValHisGlnLysThrHisThrGluGluLysProTyrGlnCysAsnGluCy 323

     618 TGGCCGAAGCTATAAACAGCGAACGTCTTTAGAGGAACATAAAGAG.... 663
         ||||:::|||:::|||:::   |||   ::::::   |||::::::
     323 sGlyLysSerPheLysAsnThrThrIlePheAsnValHisGlnArgIleH 340

     664 .....................CGCTGCCACAACTACTTGGAAAGCATGGGC 693
                              ||||||:::::      ::::::
     340 isThrGlyGluLysProPheArgCysAsnGluCysGlyLysAlaTyrArg 356

     694 CTTCCGGGCACACTGTACCCAGTCATTAAAGAAGAAACTAAGCACAGTGA 743
               :::::::|||         |||:::   :::|||   ::::::
     357 SerAsnSerSerLeuIleValHisIleArgThrHisThrGlyGluLysPr 373

     744 AATGGCAGAAGACCTGTGCAAGATAGGATCAGAGAGATCTCTCGTGCTGG 793
                    :::   |||                  :::   :::::::
     373 oTyrGluCysAsnGluCys.................GlyLysAlaPheA 384

     794 ACAGACTAGCAAGTAATGTCGCCAAACGTAAGAGCTCTATGCCTCAGAAA 843
         ::|||:::|||:::   :::      ::::::                |||
     384 snArgIleAlaAsnPheThrGluHisGlnArgIleHisThrGlyGluLys 400

     844 TTTCTTGGGGACAAGGGCCTGTCCGACACGCCCTACGACAGTGCCACGTA 893
                                       ||||||
     401 .............................ProTyr.............. 402

     894 CGAGAAGGAGAACGAAATGATGAAGTCCCACGTGATGGACCAAGCCATCA 943
             |||   ||||||      |||:::
     403 ....LysCysAsnGluCysGlyLysAlaPhe.................. 411

     944 ACAACGCCATCAACTACCTGGGGGCCGAGTCCCTGCGCCCGCTGGTGCAG 993
                 |||||||||
     412 ........IleAsnTyr............................... 414

     994 ACGCCCCCGGGCGGTTCCGAGGTGGTCCCGGTCATCAGCCCGATGTACCA 1043
```

```
                                        |||   :::
     415 ..................................SerCysLeuThrVa 419

    1044 GCTGCACAGGCGCTCGGAGGGCACCCCGCGCTCCAACCACTCGGCCCAGG 1093
          ||||||         |||           :::   :::
     419 lHisHisArgMetHisThrGlyGluLysProTyrLysCysThrGluCysG 436

    1094 ACAGCGCCGTGGAGTACCTGCTGCTGCTCTCCAAGGCCAAGTTGGTGCCC 1143
           :::|||                 ::::::   ::::::::::|||:::
     436 lyLysAla..............PheMetArgSerSerSerLeuIleIle 447

    1144 TCGGAGCGC.........GAGGCGTCCCCG......AGCAACAGCTGCCA 1178
           :::|||         |||   :::|||         |||:::|||
     448 HisGlnArgIleHisThrGluGluLysProTyrLeuCysAsnGluCysGl 464

    1179 AGACTCCACGGACACCGAGAGCAACAACGAGGAGCAGCGCAGCGGTCTTA 1228
          :::|||         :::|||
     464 yGluSerPheArgIleLysSer........................... 471

    1229 TCTACCTGACCAACCAC.................................. 1245
           :::||||||   |||
     472 ..HisLeuThrValHisGlnArgIleHisThrGlyGluLysProTyrLys 487

    1246 ......ATCGCCCGACGCGCG......CAACGCGTGTCGCTCAAGGAGGA 1283
                 ::::::||||||      :::   |||:::||||||||::
     488 CysThrAspCysGluArgAlaPheThrLysMetValAsnLeuLysGluHi 504

    1284 G..........................CACCGCGCCTACGACCTGCTGC 1306
          :                         :::::::::||||||      :
     504 sGlnLysIleHisThrGlyValLysProTyrLysCysTyrAspCysGlyL 521

    1307 GCGCCGCCTCCGAGAACTCGCAGGACGCGCTCCGCGTGGTCAGCACCAGC 1356
          :::::          :::|||         ::::::      :::   :::
     521 ysSerPheArgThrLysSerTyrLeuIleValHisGlnArgThrHisThr 537

    1357 GGGGAGCAGATGAAGGTGTACAAGTGCGAACACTGCCGGGTGCTCTTCCT 1406
          ||||||      |||   |||||||||:::::::|||:::      |||
     538 GlyGlu......LysProTyrLysCysAsnGluCysGluLysAlaPheTh 552

    1407 GGATCACGTCATGTACACCATCCACATGGGCTGCCACGGCTTCCGTGATC 1456
          :::      :::   |||:::|||:::      |||      :::   |
     552 rAsnThrSerGlnLeuThrValHisGlnArgArgHisThrGlyGluLysP 569

    1457 CTTTTGAGTGCAACATGTGCGGCTACCACAGCCAGGACCGGTACGAGTTC 1506
          ||:::::::||||||   ||||||            ::::::      |||
     569 roTyrLysCysAsnGluCysGlyLysValPheThrSerAsnSerGlyPhe 585

    1507 TCGTCGCAC 1515
          ::::::|||
     586 AsnThrHis 588

seq_name: SwissProt_40:Y441_HUMAN

seq_documentation_block:
ID   Y441_HUMAN     STANDARD;      PRT;   697 AA.
AC   O43167;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical zinc finger protein KIAA0441.
GN   KIAA0441.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98116655; PubMed=9455477;
RA   Ishikawa K.-I., Nagase T., Nakajima D., Seki N., Ohira M.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. VIII.
RT   78 new cDNA clones from brain which code for large proteins in
RT   vitro.";
RL   DNA Res. 4:307-313(1997).
CC   -!- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB007901; BAA23713.1; -.
DR   HSSP; P08047; 1SP2.
DR   InterPro; IPR000637; AT_hook.
DR   InterPro; IPR000210; BTB_POZ.
DR   InterPro; IPR000822; Znf-C2H2.
DR   Pfam; PF02178; AT_hook; 1.
DR   Pfam; PF00651; BTB; 1.
DR   Pfam; PF00096; zf-C2H2; 8.
DR   PRINTS; PR00048; ZINCFINGER.
DR   SMART; SM00384; AT_hook; 1.
DR   SMART; SM00225; BTB; 1.
DR   SMART; SM00355; ZnF_C2H2; 8.
DR   PROSITE; PS50097; BTB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 8.
KW   Hypothetical protein; Transcription regulation; DNA-binding;
KW   Zinc-finger; Metal-binding; Nuclear protein; Repeat.
FT   DOMAIN       10    133       BTB.
FT   DOMAIN      294    512       ZINC FINGERS.
FT   ZN_FING     294    316       C2H2-TYPE.
FT   ZN_FING     322    344       C2H2-TYPE.
FT   ZN_FING     350    372       C2H2-TYPE.
FT   ZN_FING     378    400       C2H2-TYPE.
FT   ZN_FING     406    428       C2H2-TYPE.
FT   ZN_FING     434    456       C2H2-TYPE.
FT   ZN_FING     462    484       C2H2-TYPE.
FT   ZN_FING     490    512       C2H2-TYPE.
SQ   SEQUENCE   697 AA;  78292 MW;  F2BD33C144626544 CRC64;


alignment_scores:
            Quality:  312.00              Length:    548
              Ratio:  1.253                 Gaps:     26
 Percent Similarity:  45.438   Percent Identity: 24.270

alignment_block:
 US-08-711-417C-165 x Y441_HUMAN   ..

 Align seg 1/1  to: Y441_HUMAN  from: 1  to: 697

      10 GACGAGGGT.........CAAGACATGTCTTTCTCATCAGGGAAGGAAAG 50
         |||:::|||         ::::::::::::   ::::::   :::|||||
     203 AspSerGlyValLeuAsnGluGlnIleAlaAlaLysGluLysGluGluSe 219

      51 CCCCCCTGTAAGCGATACTCCAGATGAGGGCGATGAGCCCATGCCGATCC 100
         |   |||:::   :::          |||      |||   ||||||:::
     219 rGluProThrCysGluProSerArgGlu......GluGluMetProValG 234

     101 CCGAGGACCTCTCC.........ACCACCTCGGGAGGACAG......CAA 135
           :::|||   :::            |||:::   ||||||      |||
     234 luLysAspGluAsnTyrAspProLysThrGluAspGlyGlnAlaSerGln 250

     136 AGCTCCAAGAGTGACAGAGTCGTGGCCAGTAATGTTAAAGTAGAGACTCA 185
         |||      |||   |||   :::      :::||||||::::::
     251 SerArgTyrSerLysArgArgIleTrpArgSerValLysLeuLysAspTy 267

     186 GAGTGATGAAGAGAATGGGCGTGCCTGTGAAATGAATGGGGAAGAATGTG 235
```

```
 267 r..................................................... 267

 236 CGGAGGATTTACGAATGCTTGATGCCTCGGGAGAGAAAATGAATGGCTCC 285
                                        |||:::   |||:::
 268 ..............................LysLeuValGlyAsp 272

 286 CACAGGGACCAAGGCAGCTCGGCTTTGTCGGGAGTTGGAGGCATTCGACT 335
     ::::::|||:::|||   ||||||            |||      :::
 273 GlnGluAspHisGly...SerAlaLysArgIleCysGlyArgArgLysAr 288

 336 TCCTAACGGA...AAACTAAAGTGTGATATCTGTGGGATCATTTGCATCG 382
      |||:::|||   :::   :::|||      ||||||   :::
 288 gProGlyGlyProGluAlaArgCysLysAspCysGlyLysValPheLysT 305

 383 GGCCCAATGTGCTCATGGTTCACAAAAGAAGCCACACTGGAGAACGGCCC 432
         :::   |||   :::|||:::|||||||||||||||||||||||||
 305 yrAsnHisPheLeuAlaIleHisGlnArgSerHisThrGlyGluArgPro 321

 433 TTCCAGTGCAATCAGTGCGGGGCCTCATTCACCCAGAAGGGCAACCTGCT 482
     |||:::||||||:::||||||   :::|||:::||||||   :::|||
 322 PheLysCysAsnGluCysGlyLysGlyPheAlaGlnLysHisSerLeuGl 338

 483 CCGGCACATCAAGCTGCATTCCGGGGAGAAGCCCTTCAAATGCCACCTCT 532
       |||   ::::::|||:::||||||:::|||:::   |||   :::|
 338 nValHisThrArgMetHisThrGlyGluArgProTyrThrCysThrValC 355

 533 GCAACTACGCC..................................... 543
     ||:::   |||
 355 ysSerLysAlaLeuThrThrLysHisSerLeuLeuGluHisMetSerLeu 371

 543 ..................................................... 543

 372 HisSerGlyGlnLysSerPheThrCysAspGlnCysGlyLysTyrPheSe 388

 543 ..................................................... 543

 388 rGlnAsnArgGlnLeuLysSerHisTyrArgValHisThrGlyHisSerL 405

 544 .................TGCCGCCGGAGG............GACGCCCTC 564
                     |||:::|||:::            :::   |||
 405 euProGluCysLysAspCysHisArgLysPheMetAspValSerGlnLeu 421

 565 ACTGGCCACCTGAGGACGCACTCCGTTGGTAAACCTCACAAATGTGGATA 614
       ||||||||||||||:::      ||||||      |||
 422 LysLysHisLeuArgThrHisThrGlyGluLysProPheThrCysGluIl 438

 615 TTGTGGCCGAAGCTATAAACAGCGAACGTCTTTAGAGGAACATAAAGAGC 664
     ||||||:::|||:::      ::::::||||||:::   |||      |
 438 eCysGlyLysSerPheThrAlaLysSerSerLeuGlnThrHis...IleA 454

 665 GCTGCCACAACTACTTGGAA......AGCATGGGCCTTCCGGGCACACTG 708
     ||   |||:::      :::      |||   |||:::   |||
 454 rgIleHisArgGlyGluLysProTyrSerCysGlyIleCysGlyLysSer 470

 709 TACCCAGTCATTAAAGAAGAAACTAAGCACAGTGAAATGGCAGAAGACCT 758
     :::         :::   :::   :::|||
 471 PheSerAspSerSerAlaLysArgArgHis..................... 480

 759 GTGCAAGATAGGATCA......................GAGAGATCTC 784
      |||   :::   :::                      |||   :::|
 481 .CysIleLeuHisThrGlyLysLysProPheSerCysProGluCysAsnL 497

 785 TCGTGCTGGACAGACTAGCAAGTAATGTCGCCAAACGTAAGAGCTCTATG 834
     ||   :::   ||||||   :::      |||      |||
 497 euGlnPheAlaArgLeuAspAsnLeuLysAlaHisLeuLysIleHisSer 513

 835 CCTCAGAAATTTCTTGGGGAC............AAGGGCCTGTCCGACAC 872
       :::|||      :::|||           :::|||   |||:::||
 514 LysGluLysHisAlaSerAspAlaSerSerIleSerGlySerSerAsnTh 530

 873 G.........................CCCTACGAC...AGTGCCA 889
     |                        |||||||:::   |||:::
 530 rGluGluValArgAsnIleLeuGlnLeuGlnProTyrGlnLeuSerThrS 547

 890 CGTACGAGAAGGAGAACGAAATGATGAAGTCCCACGTGATGGACCAAGCC 939
     ::   |||:::|||   :::::::::         |||   |||::::::
 547 erGlyGluGlnGluIleGlnLeuLeu.........ValThrAspSerVal 560

 940 ATCAACAACGCCATCAACTACCTGGGGGCCGAGTCCCTGCGCCCGCTGGT 989
        |||      ||||||::::::
 561 HisAsn......IleAsnPheMet........................... 566

 990 GCAGACGCCCCCGGGCGGTTCCGAGGTGGTCCCGGTCATCAGCCCG.... 1035
               ||||||   |||:::   :::   :::::::::
 567 ..........ProGlyProSerGlnGlyIleSerIleValThrAlaGluS 580

1036 ................................ATGTACCAGCTGCAC 1050
                                     :::      |||
 580 erSerGlnAsnMetThrAlaAspGlnAlaAlaAsnLeuThrLeuLeuThr 596

1051 AGGCGCTCGGAGGGCACCCCGCGCTCCAACCACTCGGCCCAGGACAGCGC 1100
     ::::::   |||         :::         |||||||||::::::
 597 GlnGlnProGluGlnLeuGlnAsnLeuIleLeuSerAlaGlnGlnGluGl 613

1101 CGTGGAGTACCTG......CTGCTGCTCTCCAAGGCCAAGTTGGTGCCCT 1144
      :::|||::::::      |||   :::   ::::::::::::   ||||
 613 nThrGluHisIleGlnSerLeuAsnMetIleGluSerGlnMetGlyProS 630

1145 CGGAGCGCGAGGC.............GTCCCCGAGCAACAGCTGCCA 1178
     ||:::   |||                :::   ::::::   |||
 630 erGlnThrGluProValHisValIleThrLeuSerLysGluThrLeuGlu 646

1179 AGACTCCACGGACACCGAGAGCAACAACGAGGAGCAGCGCAGCGGTCTTA 1228
     :::||||||:::|||:::|||||| :::|||||||              :
 647 HisLeuHisAlaHisGlnGluGln.ThrGluGlu.............L 658

1229 TCTACCTGACCAACCACATCGCCCGACGCGCGCAACGCGTGTCGCTCAAG 1278
     :::::|||::::::      :::      ||||||::::::::|||
 658 euHisLeuAlaThrSerThrSerAspProAlaGlnHisLeuGlnLeuThr 674

1279 GAGGAGCACCGCGCCTACGACCTGCTGCGCGCCGCCTCCG 1318
     :::|||    :::|||              |||||||||
 675 GlnGlu.ProGlyPro.............ProProPro 682
```

```
seq_name: SwissProt_40:HUNB_DROYA

seq_documentation_block:
ID   HUNB_DROYA     STANDARD;      PRT;   759 AA.
AC   O62541;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hunchback protein.
GN   HB.
OS   Drosophila yakuba (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7245;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Tautz D., Nigro L.;
RT   "Microevolutionary divergence pattern of the segmentation gene
RT   hunchback in Drosophila.";
RL   Mol. Biol. Evol. 15:1403-1411(1998).
CC   -!- FUNCTION: GAP CLASS SEGMENTATION PROTEIN THAT CONTROLS DEVELOPMENT
CC       OF HEAD STRUCTURES (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE HUNCHBACK FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AJ005376; CAA06506.1; -.
DR   FlyBase; FBgn0022824; Dyak\hb.
DR   InterPro; IPR000822; Znf-C2H2.
DR   Pfam; PF00096; zf-C2H2; 6.
DR   PRINTS; PR00048; ZINCFINGER.
DR   SMART; SM00355; ZnF_C2H2; 6.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 2.
KW   Developmental protein; Gap protein; Zinc-finger;
KW   Metal-binding; DNA-binding; Repeat; Nuclear protein.
FT   DOMAIN       61     70       POLY-GLN.
FT   DOMAIN      103    109       POLY-GLN.
FT   DOMAIN      118    123       POLY-HIS.
FT   DOMAIN      241    350       ZINC FINGERS I.
FT   DOMAIN      706    758       ZINC FINGERS II.
FT   ZN_FING     241    263       C2H2-TYPE.
FT   ZN_FING     270    292       C2H2-TYPE.
FT   ZN_FING     298    320       C2H2-TYPE.
FT   ZN_FING     326    350       C2H2-TYPE.
FT   DOMAIN      405    410       POLY-GLN.
FT   DOMAIN      525    528       POLY-GLU.
FT   DOMAIN      555    564       POLY-GLN.
FT   DOMAIN      690    695       POLY-ALA.
FT   ZN_FING     706    728       C2H2-TYPE.
FT   ZN_FING     734    758       C2H2-TYPE.
SQ   SEQUENCE   759 AA;  83287 MW;  67D37A252978D065 CRC64;


alignment_scores:
            Quality:  311.50              Length:    684
              Ratio:  1.097                Gaps:     24
 Percent Similarity:  41.520   Percent Identity: 19.737

alignment_block:
 US-08-711-417C-165 x HUNB_DROYA   ..

 Align seg 1/1  to: HUNB_DROYA  from: 1  to: 759

      31 TTCTCATCAGGGAAGGAAAGCCCCCCTGTAAGCGATACTCCAGATGAGGG 80
         |||   :::|||:::      |||      :::   ||||||   :::::
     143 PheTyrGlyGlyAsnLeuArgProSerProGlnProThrProThrSerAl 159

      81 CGATGAGCCCATGCCGATC.............................. 99
         ::::         |||:::
     159 aSerThrValAlaProValAlaValAlaThrGlySerSerGluLysLeuG 176

     100 .............CCCGAGGACCTCTCCACCACCTCGGGAGGACAGCAA 135
                       |||   |||::::::      ::::::      :::
     176 lnAlaLeuThrProProMetAspValThrProProLysSerProAlaLys 192

     136 AGCTCCAAGAGTGACAGAGTCGTGGCCAGTAATGTTAAAGTAGAGACTCA 185
         |||||||:::                  ||||||::::::   |||   ::
     193 SerSerGln.................SerAsnIleGluProGluLysGl 203

     186 GAGTGATGAAGAGAATGGGCGTGCCTGTGAAATGAATGGGGAAGAATGTG 235
         :   |||:::   :::                              :::   :
     203 uHisAspGlnMetSer.........................AsnSerS 211

     236 CGGAGGATTTACGA...ATGCTTGATGCCTCGGGAGAGAAAATGAATGGC 282
         ::||||||::::::   |||   :::::::::   :::      |||
     211 erGluAspMetLysTyrMetAlaGluSerGluAspAspAspThrAsn... 226

     283 TCCCACAGGGACCAAGGCAGCTCGGCTTTGTCGGGAGTTGGAGGCATTCG 332
                                                  |||||
     227 .............................................IleAr 228

     333 ACTTCCT............AACGGAAAACTA.........AAGTGTGATA 361
         |:::|||           :::|||||||:::         ||||||
     228 gMetProIleTyrAsnSerHisGlyLysMetLysAsnTyrLysCysLysT 245

     362 TCTGTGGGATCATTTGCATCGGGCCCAATGTGCTCATGGTTCACAAAAGA 411
           ||||||:::::::::|||            :::   :::|||   |||
     245 hrCysGlyValValAlaIleThrLysValAspPheTrpAlaHisThrArg 261

     412 AGCCAC...ACTGGAGAACGGCCCTTCCAGTGCAATCAGTGCGGGGCCTC 458
         :::|||        ::::::   :::||||||   :::|||
     262 ThrHisMetLysProAspLysIleLeuGlnCysProLysCysProPheVa 278

     459 ATTCACCCAGAAGGGCAACCTGCTCCGGCACATCAAGCTGCATTCCGGGG 508
                 |||   :::|||      |||||||:::   |||::::::::
     278 lThrGluPheLysHisHisLeuGluTyrHisIleArgLysHisLysAsnG 295

     509 AGAAGCCCTTCAAATGCCACCTCTGCAACTACGCCTGCCGCCGGAGGGAC 558
         ::||||||||||:::|||      |||:::|||:::|||   :::::::::
     295 lnLysProPheGlnCysAspLysCysSerTyrThrCysValAsnLysSer 311

     559 GCCCTCACTGGCCACCTGAGGACGCACTCCGTTGGTAAACCTCACAAATG 608
           |||:::::::|||   :::::|||||||            :::::::||
     312 MetLeuAsnSerHisArgLysSerHisSerSerValTyrGlnTyrArgCy 328

     609 TGGATATTGTGGCCGAAGCTATAAACAGCGAACGTCTTTAGAGGAACATA 658
         |:::   |||      :::   |||
     328 sAlaAspCysAspTyrAlaThrLysTyr..................... 337

     659 AAGAGCGCTGCCACAACTAC.........TTGGAAAGCATGGGC...CTT 696
                |||||||:::::         |||:::::::   |||
     338 ........CysHisSerPheLysLeuHisLeuArgLysTyrGlyHisLys 351

     697 CCGGGCACACTGTACCCAGTCATTAAAGAAGAA...ACTAAGCACAGTGA 743
         ||||||            |||:::   |||:::   |||   :::   ::
     352 ProGlyMet.........ValLeuAspGluAspGlyThrProAsnProSe 365

     744 AATGGCAGAAGACCTGTGCAAGATAGGATCAGAGAGATCTCTCGTGCTGG 793
         :::::::   |||:::         |||::::::|||:::
     365 rLeuValIleAspVal......TyrGlyThrArgArgGly.......... 376

     794 ACAGACTAGCAAGTAATGTCGCCAAACGTAAGAGCTCTATGCCTCAGAAA 843
                                  |||   |||::::::   |||
     377 ....................ProLysSerLysAsnGlyGlyProIleAla 386

     844 TTTCTTGGGGACAAGGGCCTGTCCGACACGCCCTACGACAGTGCCACGTA 893
               |||:::   :::   |||      |||      :::|||:::
     387 SerGlyGlySerGlySerGlySerArgLysProAsnValAlaAlaValAl 403

     894 CGAGAAGGAGAACGAAATGATGAAGTCCCACGTGATGGACCAAGCCATCA 943
             ::::::::::::   :::   ::::::         :::   ::::
     403 aProGlnGlnGlnGlnSerGlnProAlaGlnProAlaThrSerGlnLeuS 420

     944 ACAACGCCATCAACTACCTG............................. 963
         ::   |||::::::   :::
     420 erAlaAlaLeuGlnGlyPheProLeuValGlnSerAsnSerAlaProPro 436

     964 ................................GGGGCCGAGTCCCTGCG 980
                                        |||:::|||:::
     437 AlaAlaSerProValLeuProLeuProAlaSerProAlaLysSerValAl 453

     981 CCCGCTGGTGCAGACGCCC......CCGGGCGGTTCCGAGGTGGTCCCGG 1024
             :::   |||||||||      |||:::   :::::::::::::|||
     453 aSerValGluGlnThrProSerLeuProSerProAlaAsnLeuLeuProP 470

    1025 TCATCAGCCCGATGTACCAGCTGCACAGGCGCTCGGAGGGCACCCCGCGC 1074
           ::::::   :::   |||   :::|||:::            |||
     470 roLeuAlaSerLeuLeuGlnGlnAsnArgAsnMetAlaPhePheProTyr 486

    1075 TCCAACCACTCGGCCCAGGACAGCGCCGTGGAGTACCTGCTGCTGCTCTC 1124
            |||   :::   |||                    :::|||||||::
     487 TrpAsnLeuAsnLeuGln.....................MetLeuLeuAl 496

    1125 CAAG.................GCCAAGTTGGTGCCCTCGGAGCGCGAGG 1156
         ::::                 |||:::|||   |||      ||||||
     496 aGlnGlnGlnAlaAlaValLeuAlaGlnLeuSerProArgMetArgGlu. 512
```

```
1157 CGTCCCCGAGCAACAGCTGCCAAGACTCCACGGACACCGAGAGCAACAAC 1206
                  :::   |||:::::::::::::   :::|||:::|||
 513 .............GlnLeuGlnGlnGlnAsnGlnHisGlnSerAspAsn 524

1207 GAGGAGCAGCGCAGCGGTCTTATCTACCTGACCAACCACATC........ 1248
     ||||||:::::::::         |||      :::   :::
 525 GluGluGluGluGlnAspAspGluTyrGluArgLysSerValAspSerAl 541

1248 ..................................................  1248

 541 aMetAspLeuSerGlnGlyThrProValLysGluAspAspGlnHisGlnG 558

1249 ...........GCCCGACGCGCGCAACGCGTGTCGCTCAAGGAGGAGCAC 1287
                 ::::::         ::::::||||||   |||:::
 558 lnGlnGlnGlnProGlnGlnProLeuAlaMetAsnLeuLysValGluGlu 574

1288 CGCGCCTACGACCTG.................................  1302
     :::|||      |||
 575 GluAlaThrProLeuMetSerSerSerAsnAlaSerArgArgLysGlyAr 591

1303 ..........................CTGCGCGCCGCCTCCGAGAACT 1324
                                 ||||||:::   :::   :::|
 591 gValLeuLysLeuAspThrLeuLeuGlnLeuArgSerGluAlaMetThrS 608

1325 CGCAGGACGCGCTCCGCGTGGTCAGCACC..................  1353
     ||   :::   |||:::|||   ||||||
 608 erProGluGlnLeuLysValProSerThrProMetProThrAlaSerSer 624

1353 ..................................................  1353

 625 ProIleAlaGlyArgLysProMetProGluAspHisCysSerGlyThrSe 641

1354 ....AGCGGGGAGCAGATGAAG.........................  1371
        :::   |||:::|||:::
 641 rSerAlaAspGluSerMetGluThrAlaHisValAlaGlnAlaAsnThrS 658

1371 ..................................................  1371

 658 erAlaSerSerThrAlaSerSerSerGlyAsnSerSerAsnAlaSerSer 674

1371 ..................................................  1371

 675 AsnGlyAsnSerSerSerAsnSerSerSerAsnGlyThrSerSerAlaAl 691

1372 ....................................GTGTACAAGT 1381
                                        :::|||:::|
 691 aAlaAlaProAlaSerGlyThrProAlaAlaAlaGlyAlaIleTyrGluC 708

1382 GCGAACACTGCCGGGTGCTCTTCCTGGATCACGTCATGTACACCATCCAC 1431
     ||:::::::|||   ::::::|||   |||   |||:::||||||||||||
 708 ysLysTyrCysAspIlePhePheLysAspAlaValLeuTyrThrIleHis 724

1432 ATGGGCTGCCACGGCTTCCGTGATCCTTTTGAGTGCAACATGTGCGGCTA 1481
     ||||||   |||:::      |||   |||:::||||||||||||||||
 725 MetGlyTyrHisSerCysAspAspValPheLysCysAsnMetCysGlyGl 741

1482 CCACAGCCAGGACCGGTACGAGTTCTCGTCGCACATAACGCGAGGGGAGC 1531
           :::  .          :::      |||::::::|||:::   |
 741 uLysCysAspGlyProValGlyLeuPheValHisMetAlaArgAsnAlaH 758

1532 AC 1533
     ||
 758 is 758
```